

ELECTRONIC CIRCUITS AND TUBES

EDITORS

HARRY E. CLIFFORD AND ALEXANDER H. WING

AUTHORS

The authors of this volume were members of the Electronics Training Staff of Cruft Laboratory, Harvard University, engaged in the pre-radar training of Army and Navy officers.

E. L. CHAFFEE, A.M., PH.D.

*Director of Cruft Laboratory
Professor of Applied Physics
Harvard University*

JAMES D. COBINE, M.S., PH.D.

*Research Physicist
Research Laboratory
General Electric Company*

HARRY ROWE MIMNO, E.E., PH.D.

*Professor of Applied Physics
Harvard University*

STUART P. COOKE, PH.B., M.S.

*Research Lecturer in Physics
and Communication
Harvard University*

LLOYD W. MORRIS, PH.D.

*Professor of Physics
Louisiana State University*

R. ORIN CORNETT, PH.D.

*Vice President
Oklahoma Baptist University*

HARRY STOCKMAN, S.D.

*Chief, Communications Laboratory
Cambridge Field Station
Air Materiel Command, AAF*

SHERWOOD GITHENS, JR., M.A., PH.D.

*Physicist
Applied Physics Laboratory
The Johns Hopkins University*

G. RUSSELL TATUM, M.A., PH.D.

*Associate Director of Research
Applied Physics Laboratory
The Johns Hopkins University*

PHILIPPE E. LECORBEILLER, D.-ES-
SC.

*Lecturer in Electronics
Harvard University*

ALEXANDER H. WING, E.E., PH.D.

*Associate Professor
of Electrical Engineering
Northwestern University*

ELECTRONIC CIRCUITS AND TUBES

BY THE
ELECTRONICS TRAINING STAFF
OF THE
CRUFT LABORATORY
Harvard University

(WITH PROBLEMS ADDED)

NEW YORK AND LONDON
McGRAW-HILL BOOK COMPANY, INC.

1947

ELECTRONIC CIRCUITS AND TUBES

COPYRIGHT, 1947, BY THE
MCGRAW-HILL BOOK COMPANY, INC.
PRINTED IN THE UNITED STATES OF AMERICA

*All rights reserved. This book, or
parts thereof, may not be reproduced
in any form without permission of
the publishers.*

*To
the men and women of the
Armed Forces for whom
this work was written*

FOREWORD

This book has developed from the lecture notes of a special wartime training course given in the Graduate School of Engineering, Harvard University.

The need for highly trained officers in the new uses of electronics was definitely appreciated even in 1941. In July of that year a course for officers of the Signal Corps who were graduate electrical engineers was established at Harvard University in the Graduate School of Engineering to give intensive training in the fundamentals of electronics and high-frequency circuits. Immediately after the United States declared war, the Navy also sent officers to the pre-radar course.

The rapid expansion of the course necessitated a greatly increased instructional staff. Professors and instructors from many educational institutions were invited to Cambridge to aid the regular staff at Harvard in the field of communication engineering.

Although the war training course was distinct from the regular graduate courses in communication engineering that have been given for more than two decades in the Cruft Laboratory, it was planned originally and patterned to a considerable degree on the lines of some of the Cruft Laboratory courses. Since the scope and the method of presentation of the material in this volume are not on a graduate level, they differ somewhat from corresponding course work in the Cruft Laboratory.

The members of the wartime staff have brought to the work many excellent ideas from their teaching experience. More than thirty repetitions of the course gave opportunity to improve its character both as to lecture presentations and laboratory experiments. Considerable new material was injected into the course content. In short, this war training program presented a rare opportunity to develop and improve teaching methods in electronics.

During the progress of the course, lecture notes were prepared mainly by twelve members of the lecturing staff. It was the original intention to publish the entire lecture material in a single volume. Because the single volume would be inconveniently large and also since a portion of the manuscript was completed early, that portion has been published in a separate volume "Transmission Lines, Antennas, and Wave Guides" by Associate Profs. R. W. P. King and H. R. Mimno and Dr. A. H. Wing. The remaining and larger por-

tion of the lecture material comprises the subject matter of the present book. Logically the material of this book precedes that of the book already published; thus, the present book is Volume I of a two-volume text. Eleven of the original twelve members of the wartime lecturing staff are authors of the present book.

The text presented in this book has developed out of the intensive devotion, during a war emergency, of a group of men to a single purpose, that of imparting to the student officers in the most efficient manner a comprehensive and practical knowledge of electronics. The material of the course was, however, fundamental in nature and not exclusively applicable to wartime training. For this reason, it is hoped that the text will be as valuable for peacetime courses as it was successful in its intended purpose.

The treatment of the subject matter of this book is suitable for juniors or seniors of colleges and engineering schools, who are specializing in the study of communication engineering or in electronics. A knowledge of mathematics through calculus, and of electricity and magnetism is assumed. The book may also be found to be of value as a reference for others.

E. L. CHAFFEE

PREFACE

This book presents the basic theory of electronic tubes and electric circuits employed in conjunction with these tubes. The emphasis is on the applications in the fields of communication and electronic control. Supplementary material on mathematics and electricity and magnetism is given in the appendixes, making the book useful to those needing preparation in these subjects.

As stated in the foreword, the book has developed from the lecture notes of a special wartime training course. Many chapters of the book contain more material than was given in the course. In the preparation of the notes and the book, there was much collaboration among the authors. In this work, each chapter has been the responsibility of one or two of the authors. Chapters I, II, III, IV, XXII and Appendix C were written by L. W. Morris; Chaps. V and VI by S. P. Cooke; Chaps. VII, XIV, and part of Chap. XIX by E. L. Chaffee; Chaps. VIII, IX, XVIII and Appendix A by P. E. LeCorbeiller; Chaps. X and XI by R. O. Cornett; Chaps. XII, XV, and XXIV by H. R. Mimno; Chap. XIII mostly by G. R. Tatum, in part by H. Stockman; Chaps. XVI and XVII by J. D. Cobine; Chaps. XIX and XX by S. Githens, Jr. and A. H. Wing; Chap. XXIII by H. Stockman with the assistance of S. Githens, Jr.; Appendix B by P. E. LeCorbeiller and S. Githens, Jr.; Chap. XXI and Appendix D by A. H. Wing. This is not to be construed as a list of exclusive authorship in all cases, as there were many interchanges of ideas. There were also some modifications made in original versions in order to fit them into the general plan of the book.

The book teaches fundamental principles clearly and rigorously and contains much new material and new methods of presentation. In the treatment of a-c circuits the behavior of circuit elements with respect to nonsinusoidal waveforms is described in such a way as to develop correct concepts of resistance, inductance, and capacitance. For sinusoidal waveforms the emphasis is more on the magnitude and angle of the impedance rather than on resistance and reactance. Curves or loci are used extensively to describe circuit behavior. The treatment of resonance is rigorous and more complete than usual. There is a new presentation of the bridged-T and parallel-T networks.

In the chapter on networks and impedance matching, the subject

of equivalent four-terminal networks is discussed quite fully. Network theorems are presented, with more emphasis on Thévenin's theorem than on any of the others. The action of the tank circuit as an impedance transformer is clearly explained. The chapter on transients is fully illustrated with diagrams. The effect of the time constant in LR and CR circuits is stressed. The analysis of initial and final conditions is explained effectively.

Much hitherto unpublished material is given in Chap. VII. The method of correlating response curves of coupled circuits by means of space models and contour diagrams is unique and results in an understanding of these circuits surpassing that conveyed by the ordinary response-curve type of presentation. There is new material on the band widths of magnetically coupled circuits.

A new approach to filters is given in Chap. VIII, with added material on delay networks. A rigorous yet simplified presentation of Fourier analysis is given in Chap. IX; the effect of waveform discontinuities on the magnitude of the harmonic components is explained. The analysis of periodic phenomena is extended in Chap. XVIII to nonperiodic pulses, again by a simplified method. In view of the increased use of pulses, this material should be of considerable value to the undergraduate student.

In Chaps. X and XI the action of the vacuum tube as a device is emphasized, rather than the quantitative theory of electron emission. The explanations are given both graphically and analytically. Chapter XII contains a very complete but qualitative description of the cathode-ray tube. The effects of various combinations of magnetic and electrostatic focusing and deflection are discussed. There is an unusually clear and complete description of the modern cathode-ray oscillograph.

Chapter XIII is the longest in the book and includes a complete discussion of wide-band amplifiers. The effect of feedback on the response characteristics and the output impedance is derived. The cathode follower is treated very fully.

Extensive new material on power tubes is to be found in Chap. XIV. Contour diagrams are used to describe the behavior of the Class C stage. Many forms of coupling the load to the tube are analyzed. There is new material on methods of tuning and methods of neutralization.

Oscillators are given comprehensive but qualitative treatment in Chap. XV. The drag loop, a phenomenon commonly neglected but

of considerable importance where tuned systems are connected to oscillator tubes, is adequately explained here and in Chap. XIV. The newer types of oscillator circuits are also described.

The chapter on gas-filled tubes covers the use of these tubes as rectifiers and control elements. A complete example of a rectifier design is worked out in Chap. XVII. Voltage-regulated power supplies are covered more fully than usual.

The treatment of modulation is divided into two parts, one on principles and the other on methods. The principles are thoroughly discussed to facilitate the understanding of methods or devices, which in practice are being continually improved. The treatment of angle modulation will, it is hoped, save the student from pitfalls which have caught many in the discussion of this subject.

Linear and square-law detection are treated qualitatively and quantitatively in Chap. XXI. Oscillograms are used to show clipping in the diode detector. Rectification diagrams are given for all the important large-signal detectors. The more common discriminator circuits are qualitatively treated. This material ties in with Chap. XXIII in which radio receivers are classified by type and their operation described. The operation of mixer and converter tubes is explained with a new terminology. Receivers are not presented in this organized fashion in most text books, and this chapter fills a need for showing how circuits previously discussed separately are joined together to make a complete piece of equipment.

Chapter XXII on test instruments supplies needed and practical information. Too often the user does not appreciate just what an indicating instrument measures. New material on the behavior of vacuum-tube voltmeters is included.

A complete description of control and timing circuits is given in the last chapter. Synchronization of relaxation oscillators is discussed more clearly than usual. A new and accurate description of the blocking oscillator is given. The assembly of complete waveforming circuits from their simple elements is featured, and several circuits are worked out as examples or exercises.

Great credit is due to the members of the armed forces who were students and instructors in the Cruft courses. Their constructive criticism and hard work has been of immeasurable aid to the authors in making their work more effective. Many of the

Cruft electronics staff contributed time and effort toward the preparation of the course notes and the present book. In the preparation of the book manuscript the authors are deeply indebted to Prof. Harry E. Clifford for his valuable assistance and his helpful editing.

ALEXANDER H. WING

September, 1947.

CONTENTS

FOREWORD	vi
PREFACE.	ix

Chapter I

ALTERNATING CURRENT THEORY

I. VARYING VOLTAGES AND CURRENTS

1. Description of a Sinusoid	1
2. Addition of Sinusoids.	3
3. Peak, Average, and RMS Values.	3
4. Instantaneous and Average Power	6

II. IDEAL CIRCUIT ELEMENTS

5. Capacitor.	7
6. Resistor.	9
7. Inductor	9

III. SERIES CIRCUITS

8. Voltage and Current in a Series Combination	10
9. Kirchhoff's Voltage Law in an A-c Circuit.	12
10. Series Circuits with Dissipative Reactors	13
11. Mutual Inductance in a Series Circuit	14

IV. PARALLEL CIRCUITS

12. Voltage and Current in a Parallel Combination	14
13. Impedance of a Parallel Circuit	15
14. Equivalent Representations of Dissipative Reactors	16
15. Current and Voltage Relations in a Parallel Circuit.	18
16. Electrical Representation of an A-c Generator.	19

Chapter II

CIRCUIT RESPONSE

I. RESPONSE OF SIMPLE LR AND CR UNITS

1. Variations of Z and Y for Circuits of Constant R	21
2. Variations of Z and Y for Simple LR and CR Units	22
3. Variation of Load Voltage with Constant-current Generator.	25
4. Voltage-divider Response.	26
5. Generator of Intermediate Resistance.	27
6. Compensated Voltage Divider.	29
7. Phase-shifting Networks	30

II. SERIES RESONANCE

8. Circuit Relations at Series Resonance.	31
9. Energy and Power Relations at Series Resonance	32
10. The Frequency Response of a Series <i>LCR</i> Circuit	33
11. Calculation of Z and Y for a Range of Frequencies.	36
12. Variation of Capacitance in a Series <i>LCR</i> Circuit.	40

III. PARALLEL RESONANCE

13. Conditions at Parallel Resonance.	42
14. Response of a High- Q Parallel Circuit near Resonance.	45
15. Response of Intermediate and Low- Q Parallel Circuits	46
16. Fractional Values of Q_r	48
17. Division of Resistance between Branches in a Parallel Circuit	50

Chapter III

CIRCUIT ELEMENTS

1. The Use of Equivalent Circuits to Describe the Electrical Properties of Circuit Elements.	52
2. Electrical Representation of an Isolated Inductor or Resistor	53
3. Frequency Dependence of Inductor Characteristics.	54
4. Frequency Dependence of Resistor Characteristics.	58
5. Skin Effect	62
6. Electrical Representation of a Capacitor	65
7. Capacitor Characteristics.	66
8. Types of Capacitors	67
9. Electrolytic Capacitors.	69
10. Effects of Moisture.	71
11. Shielding	72
12. Relation of Electrical Measurements to the Circuit Representation of an Element	76

Chapter IV

MEASUREMENT OF CIRCUIT ELEMENTS

1. Theory of A-c Bridge.	79
2. Simple Types of Bridges	80
3. Application of Bridge Methods	84
4. Precautions at High Frequencies.	87
5. Bridged T and Parallel-T Measurements	90
6. Single-circuit Resonance Methods	93
7. Evaluation of Equivalent Circuit Constants from Measurements. . . .	99

Chapter V

NETWORKS AND IMPEDANCE MATCHING

1. Introduction.	98
2. Network Notation and Equations	99

3. Inductive Coupling.	104
4. Equivalent Networks.	106
5. Network Theorems.	109
6. Impedance Transformation	113
7. Radio-frequency Transforming and Matching Networks.	114
8. Mechanical Analogies in the Impedance-matching Problem	124

Chapter VI

TRANSIENTS

1. Introduction.	126
2. Resistance-Inductance Circuit.	126
3. Capacitance-Resistance Circuit	138
4. Applications of CR Circuits.	145
5. Series LCR Circuit.	153
6. Mechanical Analogues of the LCR Circuit.	165
7. Square-wave Testing.	167

Chapter VII

COUPLED CIRCUITS

1. Introduction.	176
2. Coefficient of Coupling	177

I. TRANSIENT OSCILLATIONS

3. General Considerations.	178
4. Frequencies of Transient Oscillations.	179
5. Derivation of Frequencies of Free Oscillation of Magnetically Coupled Circuits.	181

II. FORCED OSCILLATIONS

6. General Considerations.	182
7. Equivalent Impedance of Coupled Circuits	183
8. Magnitude of Secondary Current.	184
9. Models for Secondary Current.	186
10. Conditions for Partial Resonance.	189
11. Conditions for $ I_2 _{\max}$	194
12. Critical Coupling.	195
13. Value of $ I_2 _{\max}$	195
14. Recapitulation.	196
15. Variation of $ I_2 $ with ω	197
16. Resonance Curve when $f_1 = f_2$	199
17. Value of $ I_2 / I_2 _{\max}$ for Optimum Coupling	202
18. Band Width for Optimum Coupling	202
19. Band-width Curves for k_e and k_o	203
20. Parallel-fed Coupled Systems	204
21. Power Transfer in Coupled Circuits	206

Chapter VIII

FILTERS

1. Definitions	209
2. Low-pass Filter	209
3. Ladder-type Low-pass Filter.	212
4. Characteristic Impedance of Low-pass Filter.	213
5. Attenuation and Phase Lag of the Low-pass Filter	216
6. Low-pass Filter of Several Identical Sections.	219
7. Low-pass Filter as Delay Network.	219
8. High-pass Filter.	220
9. Properties of Two-terminal Nondissipative Networks.	222
10. Inverse Two-terminal Networks	223
11. Constant- k Filters	224
12. Attenuation Constant and Phase Lag of Constant- k Filters	225
13. Band-pass Filter.	227
14. Band-stop Filters	229
15. Π -section Filters	229
16. Lattice Structures	230
17. Effect of Losses	231
18. Other Types of Filter Circuits.	231
19. Low-pass Composite Filter	232
20. Applications.	234

Chapter IX

FOURIER ANALYSIS

1. Periodic Function	236
2. Object of Fourier Analysis	236
3. Odd and Even Components (C and S Components)	237
4. Analysis of an Even Function $C(t)$	239
5. Odd Function $S(t)$	240
6. Complete Fourier Analysis	240
7. Amplitude and Phase of Each Harmonic Component.	241
8. Quarter-wave Symmetries.	243
9. Absence of Even Harmonics.	244
10. Examples of Fourier Developments.	244
11. Waves Having Discontinuities.	248
12. Use of Fourier Series in Transmission Problems	249
13. Example of an Integrator.	251
14. Example of Negligible Distortion	252
15. Conditions of Nondistortion.	253
16. A Numerical Method of Fourier Analysis.	255
17. Analysis of a Simple Specific Case	258

Chapter X

ELECTRON EMISSION AND THE DIODE

1. Thermionic Emission.	260
2. Other Types of Emission	264

3. Current-Voltage Characteristics of Diodes (Static Characteristics) . . .	265
4. Space Charge	268
5. Potential Distribution in a Plane Diode.	269

Chapter XI

MULTIELEMENT TUBES

1. The Control Grid	272
2. Characteristic Curves.	273
3. Operating Point, Path of Operation	275
4. Power Relations in the Plate Circuit	279
5. Cathode Bias	280
6. Nonlinear Operation	281
7. Vacuum-tube Symbols	282
8. Tube Coefficients	283
9. Equivalent Plate Circuit	286
10. Phase Relations at Low Frequency.	288
11. Constant-current Form of the Equivalent Plate Circuit.	289
12. Interelectrode Capacitances and Input Impedance	291
13. Input Impedance with Resistive Load	293
14. Input Impedance, General Case	293
15. Tetrodes and Pentodes	295
16. Special and Multipurpose Tubes.	301

Chapter XII

CATHODE-RAY TUBES

1. Introduction	304
2. Focusing, Electric-lens Method	305
3. Intensity Control	307
4. Deflection of the Beam by an Electric Field.	308
5. Applications of Cathode-ray Tubes.	311
6. Deflection of the Beam by a Magnetic Field.	314
7. Focusing, Magnetic-lens Method	318
8. Cathode-ray Oscillograph.	320

Chapter XIII

AMPLIFIERS—CLASS A AND CLASS B

1. The Vacuum Tube as an Amplifier.	325
2. Classification of Amplifiers	326
3. Distortion in Amplifiers.	328
4. Cascading of Amplifier Stages	332
5. Direct-coupled Amplifiers.	332
6. Resistance-coupled Amplifiers	334
7. Polarizing Potentials in a Resistance-coupled Stage.	335
8. Equivalent Plate Circuits of Resistance-coupled Amplifiers	336

9. Voltage Amplification at Mid-frequencies	337
10. Voltage Amplification at Low Frequencies	339
11. Voltage Amplification at High Frequencies	343
12. Special Case, $R_b \ll r_p$ and $R_b \ll R_c$	346
13. Universal Amplification Curve for a Resistance-coupled Stage	346
14. Choice of Circuit Elements	248
15. Effects of Variations in R_b and \bar{E}_{bb}	349
16. Impedance-coupled Amplifiers	350
17. Wide-band (Video) Amplifiers	350
18. Shunt-peaking Compensation at High Frequencies	352
19. Figure of Merit	356
20. Example of High-frequency Design Procedure of a Single Stage	357
21. Low-frequency Compensation	359
22. Untuned Transformer-coupled Amplifiers	360
23. Single-tube Class A Audio-frequency Power Amplifiers	371
24. Push-pull Amplifiers	382
25. Phase Inverters	397
26. Tuned Radio-frequency Amplifiers	398
27. Tuned-secondary Transformer-coupled Amplifiers	402
28. Tuned-primary Tuned-secondary Transformer-coupled Amplifiers	405
29. Feedback in Amplifiers	410
30. Current- and Voltage-controlled Feedback	413
31. Cathode Degeneration	416
32. The Cathode Follower	417
33. Motorboating	421

Chapter XIV

POWER TUBES

1. Definition of Power Tubes	423
2. Structural Features of Power Tubes	423
3. Heating of Filament	426
4. High-frequency Tubes	427
5. Classification	428
6. General Principle of Operation of a Power-converter Tube	429
7. Quantitative Relationships in Narrow-band Operation	433
8. Testing and Operation of Power Tubes	437
9. Static Characteristic Curves of Power Tubes	438
10. Path of Operation	440
11. Calculation of the Operation Condition	443
12. Selection of Path and Contour Diagrams for Fixed Q Point	445
13. Neutralization	450
14. Practical Circuits and Tuning	454
15. Calculation of Output Circuits of a Power Amplifier	462
16. Class C Amplifier	464
17. Class B Amplifier	464
18. Grid-modulated Power Amplifier	466
19. Plate-modulated Power Amplifier	468
20. Oscillators	471

CONTENTS

xix

21. Secondary Emission from the Grid	477
22. Wide-band Power Converters, Nonsinusoidal Operation	478
23. Frequency Multipliers	479

Chapter XV

OSCILLATORS

1. Introduction	482
2. Regenerative Amplifier	482
3. Oscillator Circuits	486
4. Practical Considerations, Plate-voltage Supply	489
5. Practical Considerations, Grid-voltage Supply	491
6. Practical Considerations, Power Output	493
7. Coupled-circuit Oscillators, Drag Loop	496
8. Crystal Oscillator	499
9. Other Electromechanical Oscillators, Magnetostriction	505
10. Beat-frequency Oscillator	507
11. <i>RC</i> Oscillators	509
12. Wien Bridge Type of <i>RC</i> Oscillator	511
13. Negative-resistance Oscillators	514

Chapter XVI

GAS-FILLED TUBES

1. Electrical Discharges in Gases	521
2. Glow Tube	523
3. Gaseous Thermionic Diode	525
4. Arc Rectifiers	526
5. Gaseous Thermionic Triode, or Thyatron	529
6. Current Ratings of Arc Tubes	532

Chapter XVII

RECTIFIERS AND POWER SUPPLIES

1. Introduction	534
2. Half-wave Rectifier Circuits	536
3. Full-wave Rectifier	546
4. Voltage-doubling Rectifier Circuits	549
5. Polyphase Rectifiers	553
6. Harmonic Content of Rectified Voltages	556
7. Transformer Utilization Factor	559
8. Power Filters	562
9. Critical Inductance	565
10. Design of a Power Supply	567
11. Capacitor-input Filters	570
12. Resistance-capacitance Filter	570
13. Voltage-regulated Power Supplies	571

Chapter XVIII

SIGNAL ANALYSIS

1. Statement of the Problem. 579

FOURIER ANALYSIS OF SYMMETRICAL PULSE

2. Transmission of Nonperiodic Voltage. 580
3. Rectangular Pulse 581
4. Symmetrical Voltage Pulse 585
5. Nominal Signal Duration and Nominal Cutoff. 586
6. Spectra of Various Symmetrical Pulses 587
7. Frequency Band for Transmission of Signal, Energy Spectrum. 589
8. Character of Pulse and Width of Its Spectrum. 590
9. Effect of Decreasing Duration of Pulse 590
10. Smoothing a Signal. 591
11. Unit Function and Unit Pulse. 592
12. Analysis of Antisymmetrical Pulse. 594
13. Spectra of Various Antisymmetrical Pulses 595
14. Signal Consisting of a Finite Number of Sinusoidal Cycles 596
15. General Case 598
16. Morse Dot, or Pulse 599
17. Fourier Transform. 600

DISTORTION OF SIGNAL GOING THROUGH A TRANSDUCER

18. Fourier Analysis of Output Signal 601
19. Deprivation of Output Signal from Input Signal. 602
20. Response of Network to Unit Pulse 602
21. Interdependence of Attenuation and Phase Characteristics of Network 604
22. Conditions for No Distortion 605
23. Different Types of Distortion 606
24. Effect of Linear Distortion on Symmetrical Pulse 608
25. Method for Obtaining Shape of Output Signal When Distortion Is Slight 609

Chapter XIX

PRINCIPLES OF MODULATION

1. The Modulating Signal and Types of Modulation 612
2. Amplitude Modulation 615
3. Side Frequencies and Side Bands Due to Amplitude Modulation. 617
4. Power during Amplitude Modulation. 619
5. Representation of an Amplitude-modulated Wave by Rotating Vectors 621
6. Angle Modulation 623
7. Phase Modulation 625
8. Frequency Modulation 630
9. Frequency Modulation from Phase Modulation 638

10. Alteration of Modulated Wave by a Correctly Tuned Circuit	640
11. Distortion of a Modulated Wave by an Off-tune Circuit	645
12. Distortion of a Modulated Wave from Other Causes	645
13. Spectra of Typical Signals.	645

Chapter XX

METHODS OF MODULATION

I. METHODS OF AMPLITUDE MODULATION

1. General.	648
2. Plate-modulated Class <i>C</i> Amplifier.	649
3. Grid Modulation of a Class <i>C</i> Amplifier.	655
4. Other Methods of Modulation of a Class <i>C</i> Amplifier.	656
5. Amplitude-modulated Oscillator	657
6. Production of an Amplitude-modulated Wave by Use of a Nonlinear Impedance	657
7. Triode Square-law Modulator	659
8. Balanced Modulator	660

II. METHODS OF ANGLE MODULATION

9. General.	661
10. Reactance-tube Modulation of an Oscillator.	664
11. Production of Frequency Modulation by Phase Modulation.	668

Chapter XXI

DETECTION

1. Introduction	674
2. Square-law and Linear Detection.	675

LARGE-SIGNAL (LINEAR) DIODE DETECTION

3. Large-signal Rectification.	675
4. Diode with <i>RC</i> Load.	678
5. Rectification Diagram	681
6. Graphical Solution.	682
7. Distortion and Maximum Allowable Modulation.	686
8. Automatic Volume Control	690
9. R-f Filtering	691
10. Analytical Solution, Equivalent Circuit.	692

SMALL-SIGNAL (SQUARE-LAW) DIODE DETECTION

11. Difference between Linear and Square-law Diode Detection.	693
12. Square-law Rectification	694
13. Square-law Detection, By-passed Resistive Load.	696
14. Equivalent Circuit.	698

15. Square-law Detection, Amplitude-modulated Signal	699
16. Detector-circuit Impedances	703
17. Selective Fading	704
18. Carrier and Side-band Suppression	704

DETECTION IN MULTIELEMENT TUBES

19. Grid-circuit Detection	705
20. Plate-circuit Detection	707

DETECTION OF FREQUENCY-MODULATED SIGNALS

21. Tuned Circuit as a Discriminator	712
22. Double-tuned Center-tapped Discriminator Circuit	713

Chapter XXII

TEST INSTRUMENTS

1. Direct-current Ammeters	717
2. Direct-current Voltmeters	717
3. Ohmmeters	718
4. Tube Testers	721
5. A-c Voltmeters and Ammeters	721

VACUUM-TUBE VOLTMETERS

6. D-c Measurements	724
7. A-c Measurements	727
8. Characteristics of Typical Commercial Vacuum-tube Voltmeters	732
9. Interpretation of Vacuum-tube Voltmeter Reading for Nonsinusoidal Waves	734
10. Signal Generators	737

Chapter XXIII

RADIO RECEIVERS

1. General	741
----------------------	-----

TYPES NOT EMPLOYING FREQUENCY CONVERSION

2. Regenerative Detector Receiver	741
3. Superregenerative Detector Receiver	744
4. Tuned-radio-frequency Receivers	748

SUPERHETERODYNE RECEIVERS

5. The Superheterodyne Principle	750
6. Superheterodyne Frequency Conversion	753
7. Superheterodyne Converter Circuits	758

CONTENTS

xxiii

8. The Intermediate-frequency Amplifier	761
9. Image Interference and Its Reduction	762
10. Preamplifier, Detector, Low-frequency Amplifier.	763
11. Automatic Volume Control	764
12. Automatic Frequency Control.	767

SHORT-WAVE RECEIVERS

13. General.	768
14. Short-wave Superheterodyne	769
15. Frequency-modulation Receivers	771

DISTURBANCE TO RADIO RECEPTION

16. Fluctuation Noise	774
17. Impulse Noise.	775
18. Interference.	776
19. Other Disturbances	778
20. Radiation from Receivers.	778

RECEIVER MEASUREMENTS AND SERVICING

21. Measurement of Sensitivity, Selectivity, and Fidelity.	778
22. Measurement of Signal-to-noise Ratio	779
23. Fundamentals of Servicing and Alignment	780

Chapter XXIV

TIMING CIRCUITS

1. General Remarks	783
2. Simple Relaxation Oscillator.	784
3. Frequency Adjustment.	787
4. Synchronization	788
5. Oscilloscope Synchronization	791
6. Pulse Synchronization	791
7. Frequency Division	792
8. Frequency Multiplication.	794
9. Frequency Addition and Subtraction.	795
10. Thyatron Sweep Circuits.	795
11. Improvement of Linearity.	797
12. Thyatron Pulse Circuits	799
13. High-vacuum Discharge Tubes	802
14. Blocking Oscillator.	804
15. Saw-tooth Voltage System	809
16. Saw-tooth Current Wave	810
17. Clippers	813
18. Rectangular-wave Generators	819
19. Damping Tubes	821

20. Clamping Tubes.	823
21. "Differentiators," or "Peakers".	827
22. "Integrators".	830
23. Saturable-core Reactors.	831
24. Keyers, or Electronic Switches.	835
25. "Mixer" Tubes, or Adding Tubes	837
26. Simple Applications and Combinations	839
27. Trigger Circuits	842
28. Multivibrators.	855

APPENDIXES

A. REVIEW OF MATHEMATICS	865
B. BRIEF REVIEW OF ELECTRICITY AND MAGNETISM	906
C. DESCRIPTION OF A DIRECT-CURRENT SOURCE IN TERMS OF ELECTRICAL CONSTANTS	925
D. SYMBOLS AND ABBREVIATIONS	929
PROBLEMS	935
SUBJECT INDEX	981

ELECTRONIC CIRCUITS AND TUBES

CHAPTER I

ALTERNATING CURRENT THEORY

The principles of direct-current circuit theory correctly describe the relationships between the *instantaneous* values of *varying* currents and voltages in electrical circuits. The extension of these direct-current concepts to alternating-current theory is summarized in this chapter.

I. VARYING VOLTAGES AND CURRENTS

Periodic variations are variations that repeat themselves indefinitely in time. Of these, sinusoidal variations are of particular importance because other types of periodic variation can be represented as a sum of sinusoids of different frequencies. The application of a periodic voltage in a circuit containing constant resistance, inductance, and capacitance produces, in general, currents and voltages of different waveforms. A sinusoidal or simple harmonic voltage, however, applied to a circuit of linear elements, produces currents and voltages of the same waveform and period. The purpose of a-c analysis is to relate these sinusoidal currents and voltages in terms of the circuit constants.

1. Description of a Sinusoid.—The instantaneous values of the sinusoidal voltage e and current i in Fig. 1.1*a* are related to time by

$$e = |\hat{E}| \sin \omega(t + t_0) \quad \text{and} \quad i = |\hat{I}| \sin \omega t \quad (1.1)$$

where $|\hat{E}|$ and $|\hat{I}|$ are the amplitude magnitudes of voltage and current and ω is the angular velocity or angular frequency in radians per second, related to the frequency f in cycles per second and to the period T by $\omega = 2\pi f = 2\pi/T$. The expression for e includes t_0 , which describes the shift of the sine wave of voltage toward

earlier time. The current sinusoid has no initial phase angle in (1.1) owing to the choice of zero time. The current in this example is convenient for phase reference, and t_0 is the time interval by which the voltage variations lead the corresponding current variations.

These sinusoidal variations can be described as the vertical projections of vectors rotating counterclockwise with a constant angular velocity ω , with lengths equal to the amplitude magnitudes $|\hat{E}|$ and $|\hat{I}|$, and with positions at zero time as shown in Fig. 1.1b. The angle θ_0 measures the fractional part of a period by which the voltage sinusoid leads the current sinusoid; θ_0 is equal to t_0/T times 2π radians. The two vectors rotate with the same angular velocity, their magnitude and phase relations remaining fixed.

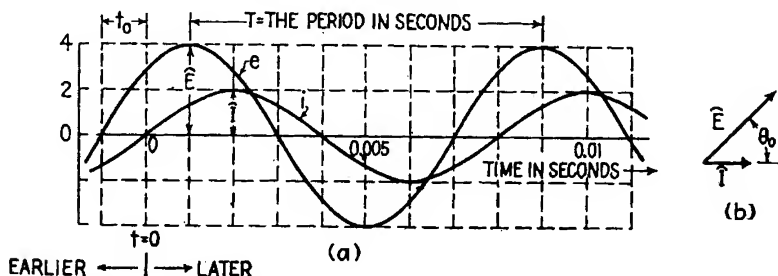


FIG. 1.1.—Sinusoidal voltage and current variations of different amplitudes and phases. For the curve marked e , $\hat{E} = 4$, $T = 0.008$ sec, $t_0 = 0.001$ sec, $f = 125$ cycles/sec, $\omega = 78.5$ radians/sec.

It is possible to describe the representative vectors at any time by the use of complex numbers in polar form,

$$\hat{E} = |\hat{E}|/\underline{\theta_E} \quad \text{and} \quad \hat{I} = |\hat{I}|/\underline{\theta_I} \quad (1.2)$$

The quantity \hat{E} is complex, having a magnitude $|\hat{E}|$ and an angle θ_E with respect to some reference vector; \hat{E} may be considered to give the magnitude and phase of the sinusoidal voltage or of the vector that describes it. The ratio of complex current and complex voltage defines two important complex constants Z and Y , either of which completely describes the magnitude and phase relations of the sinusoids. In polar form,

$$\begin{aligned} \text{Impedance} \\ Z = |\underline{Z}|/\underline{\theta_Z} &= \frac{\hat{E}}{\hat{I}} = \frac{|\hat{E}|}{|\hat{I}|} / \underline{\theta_E - \theta_I} \text{ ohms} \\ \text{Admittance} \\ Y = |\underline{Y}|/\underline{\theta_Y} &= \frac{|\hat{I}|}{|\hat{E}|} / \underline{\theta_I - \theta_E} \text{ mhos} \end{aligned} \quad (1.3)$$

The magnitude of the impedance Z in ohms is the ratio of the magnitude of the voltage amplitude in volts to the magnitude of the current amplitude in amperes; θ_z is the phase angle of the voltage with respect to the current, which for these vectors is the angle θ_0 of Fig. 1.1b. The magnitude of the ratio of the current to the voltage is given by the magnitude of Y , and θ_y is the phase angle of the current with respect to the voltage.

$$Y = \frac{1}{Z} \quad \text{and} \quad \theta_y = -\theta_z \quad (1.4)$$

The voltage across an element and the current through it establish its impedance and admittance without reference to its physical construction. A voltage sinusoid may be compared with another voltage sinusoid of the same frequency, or a current compared with a current, by taking the ratio of the complex numbers used to describe them; the resulting ratio is a complex numeric. Z and Y have the dimensions of ohms and mhos.

2. Addition of Sinusoids.—The sum of two sinusoids may be obtained by adding algebraically the instantaneous values of the sinusoids for all values of time. If the frequencies are the same, the resultant is a sinusoid of the same frequency, of an amplitude different in general from that of either component sinusoid, and of intermediate phase.

If the representative vectors of the two sinusoids are added, the projection of their vector sum describes the instantaneous value of the resultant sinusoid. If only the magnitude and phase relations are desired, the vectors may be added graphically, or their complex expressions may be added to obtain the amplitude and phase of their sum. It follows that the sum of any number of sinusoids of the same frequency is a sinusoid of that frequency. Conversely, any sinusoid of a given frequency may be represented as the sum of component sinusoids of that frequency.

3. Peak, Average, and RMS Values.—Various characteristics of periodic currents or voltages are used for their measurement. In some cases it is the peak value of a voltage, or its maximum excursion in potential, that is of interest. This is indicated by the symbol \hat{E} . For a sinusoid \hat{E} has a magnitude equal to that of the amplitude.

In some cases it is the average value of a current that is desired. This is determined by its average rate of flow of charge and is equal to the direct, or steady, current that transports the same amount

of charge during the same interval of time as the varying current. The average, or steady, current is designated by $(i)_{av}$ or simply \bar{I} . For any waveform, \bar{I} is equal to the average value of the instantaneous current evaluated over one period. The steady current of Fig. 3.1a transports a charge $q = (i)_{av} t$ in the time t , q being proportional to the area of the rectangle. For the repeated rectangular pulse of Fig. 3.1b, the average current $(i)_{av}$ is equal to the area under the pulse divided by the period. In the triangular waveform of Fig. 3.1c, the current reverses and the *net* area is divided by the period to obtain the average current. The area beneath each parabola, Fig. 3.1d, equals one-third the base times the height. The sinusoid, Fig. 3.1e, has an average value zero since

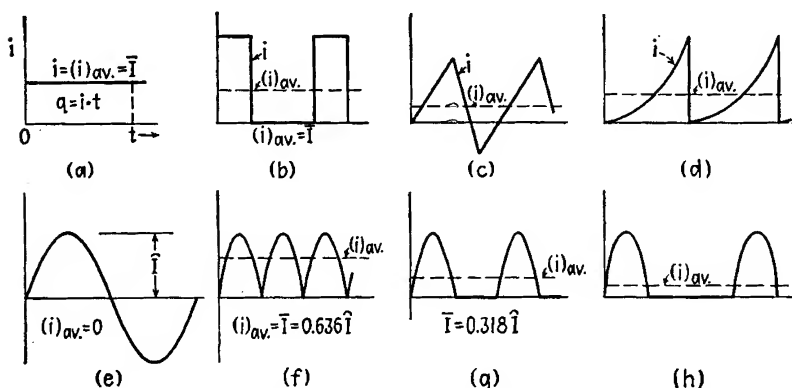


FIG. 3.1.—Average value of various currents.

its positive and negative areas are equal. In Fig. 3.1f the full-wave rectified current has an average value $\bar{I} = (2/\pi)|\hat{I}| = 0.637|\hat{I}|$ since the area beneath a half sinusoid equals $2/\pi$ times the base times the height. The average value of the half-wave rectified current, Fig. 3.1g, is $(1/\pi)|\hat{I}| = 0.318|\hat{I}|$. Where the spacing between identical positive current pulses is increased, Fig. 3.1h, the average value decreases but still remains equal to the area under each loop divided by the period. If the pulses are the upper parts (near the peaks) of a sinusoid, they closely resemble parabolas, and the area under each pulse is approximately two-thirds the base times the height.

The effective, or rms (root-mean-square), value of a varying current is equal to the direct current that, flowing in the same resistance, produces the same average power dissipation. The instantaneous power dissipation p and the average power dissi-

pation P in a resistor are

$$p = Ri^2 \quad \text{and} \quad P = R(i^2)_{\text{av}} \quad (3.1)$$

where $(i^2)_{\text{av}}$ is the average value of i^2 . If $(i^2)_{\text{av}}$ is known, the magnitude of the rms current, $|I_{\text{rms}}|$, can be evaluated. The average power dissipated is

$$P = R|I_{\text{rms}}|^2 \quad (3.2)$$

and therefore

$$|I_{\text{rms}}| = \sqrt{(i^2)_{\text{av}}} \quad (3.3)$$

Hence the root-mean-square current is the square root of the average value of the squared current.

For a sinusoidal current, $i = |\hat{I}| \sin \omega t$, the value of i^2 can be written

$$i^2 = |\hat{I}|^2 \sin^2 \omega t = |\hat{I}|^2 \left(\frac{1 - \cos 2\omega t}{2} \right) \quad (3.4)$$

As the average value of the cosine term is zero, the average value of i^2 reduces to

$$(i^2)_{\text{av}} = \frac{|\hat{I}|^2}{2} \quad \text{or} \quad I_{\text{rms}} = \frac{|\hat{I}|}{\sqrt{2}} = 0.707|\hat{I}| \quad (3.5)$$

In Fig. 3.2a the instantaneous values of i^2 from (3.4) are shown with the corresponding values of i . The curve for i^2 has sinusoidal

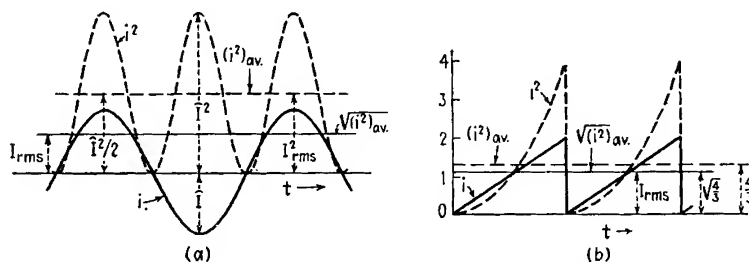


FIG. 3.2.—Root-mean-square (rms) values of sinusoidal and nonsinusoidal waves. variations of double the frequency of the current i displaced vertically and has an average value $|\hat{I}|^2/2$. The rms value of the current is indicated.

The rms value of a nonsinusoidal waveform can be obtained from a similar construction. In Fig. 3.2b the triangular waveform of i yields a series of parabolic segments for i^2 whose average value is one-third their height. The root-mean-square current is the square root of the average i^2 , as indicated in the figure.

A periodic nonsinusoidal current can be described as the sum of a direct current plus a number of sinusoidal currents whose frequencies are integral multiples of a fundamental frequency, Chap. IX. The power dissipation caused by such a current in a constant resistance R is equal to the sum of the dissipations of the separate current components calculated independently. If the d-c component is denoted by $|\bar{I}|$, and the rms values (magnitudes) of the fundamental, second harmonic, third harmonic, . . . , sinusoidal a-c components are denoted by $|I_1|$, $|I_2|$, $|I_3|$, . . . , the rms value $|I|$ of the periodic nonsinusoidal current is given by

$$|I| = \sqrt{|\bar{I}|^2 + |I_1|^2 + |I_2|^2 + |I_3|^2 + \dots} \quad (3.6)$$

The average and rms values of a periodic voltage are equal to the direct voltages that, applied across a constant resistance, produce the same average current and the same heating effect. These values may be obtained in the same manner as for a periodic current. It follows from these definitions that in a purely resistive circuit Ohm's law holds for currents and voltages of any waveform when expressed in either average or rms values.

4. Instantaneous and Average Power.—The instantaneous power delivered to a device is the product of the instantaneous voltage across it and the instantaneous current through it, for currents and voltages of any waveform. If this product varies with time, its average value gives the average power.

Let the voltage and current variations be sinusoidal according to

$$e = |\hat{E}| \sin \omega t \quad \text{and} \quad i = |\hat{I}| \sin (\omega t - \alpha) \quad (4.1)$$

The current lags the voltage by an angle α ; the instantaneous power p is

$$p = ei = |\hat{E}||\hat{I}| \sin \omega t \sin (\omega t - \alpha) \quad (4.2)$$

Using the identity $\sin a \sin b = [\cos (a - b) - \cos (a + b)]/2$ with $a = \omega t$ and $b = \omega t - \alpha$,

$$p = \frac{|\hat{E}\hat{I}|}{2} \cos \alpha - \frac{|\hat{E}\hat{I}|}{2} \cos (2\omega t - \alpha) \quad (4.3)$$

Since α is a constant, the power developed consists of a constant term $(|\hat{E}\hat{I}|/2) \cos \alpha$ and a sinusoidally varying term of double frequency whose average value is zero. The average power P is

$$P = \frac{|\hat{E}\hat{I}|}{2} \cos \alpha = \frac{|\hat{E}|}{\sqrt{2}} \frac{|\hat{I}|}{\sqrt{2}} \cos \alpha = |E||I| \cos \alpha \quad (4.4)$$

where $|E|$ and $|I|$ are the magnitudes of the effective, or rms, values. The term $\cos \alpha$ is defined as the power factor, whose value ranges between zero and unity.

The variation of the instantaneous power p with time is illustrated in Fig. 4.1 as related to the corresponding instantaneous values of current and voltage. In Fig. 4.1a the value of α is 90° (corresponding to the product of a sine and a cosine term), the instantaneous power is a sinusoid, of frequency double that of the current and voltage and of zero average value. In Fig. 4.1b where

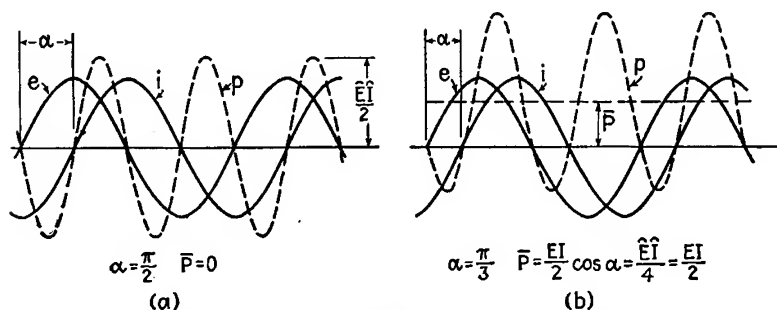


FIG. 4.1.—Instantaneous current, voltage, and power for a phase difference of (a) 90° and (b) less than 90° .

$\alpha = 60^\circ$, the sinusoidal curve for p is lifted upward by an amount \bar{P} and displaced in time as compared with Fig. 4.1a. For zero phase difference between voltage and current ($\alpha = 0^\circ$), the variation of p is similar in form to that of i^2 in Fig. 3.2a.

II. IDEAL CIRCUIT ELEMENTS

The current-voltage relations for capacitors, resistors, and inductors are summarized here, it being assumed that the circuit elements are ideal or that it is possible to describe them in terms of capacitance alone, resistance alone, or inductance alone.

5. Capacitor.—A simple form of capacitor consists of conducting films or plates separated by an insulating medium whose thickness is small compared with the surface dimensions of the films or plates. As employed in ordinary circuits, a capacitor has equal and opposite charges on the plates when there is a potential difference between them. If the instantaneous charge q is changing with time, equal charging currents flow to one plate and away from the other. The

varying charge q , the resulting instantaneous potential difference e_c (in the direction opposite to the positive flow of current), and the current i are related to each other and to the capacitance by the following expressions, which hold for any type of variation:

$$q = C e_c \quad e_c = \frac{q}{C} \quad \text{and} \quad i = \frac{dq}{dt} = C \frac{de_c}{dt} \quad (5.1)$$

The current depends upon the time rate of change or the slope of the voltage-time curve, Fig. 5.1b. In general, the current waveform differs from that of the voltage, but a voltage sinusoid yields a current sinusoid advanced in phase by one-quarter period.

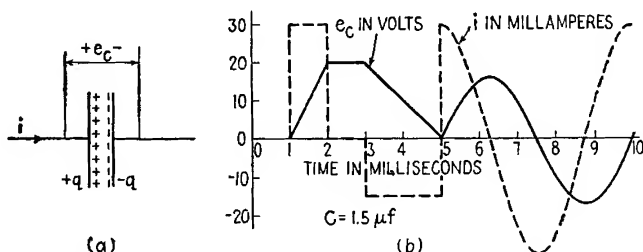


FIG. 5.1.—Ideal capacitor and its instantaneous current-voltage relations.

From (5.1) the values of e_c and i for a sinusoidal variation $q = |\hat{Q}| \sin \omega t$ are

$$e_c = \frac{|\hat{Q}|}{C} \sin \omega t \quad i = |\hat{Q}| \omega \cos \omega t \quad (5.2)$$

or

$$e_c = |\hat{E}| \sin \omega t \quad i = |\hat{I}| \sin \left(\omega t + \frac{\pi}{2} \right) \quad (5.3)$$

where $|\hat{E}| = |\hat{Q}|/C$, $|\hat{I}| = |\hat{Q}| \omega$, and $\cos \omega t$ is replaced by

$$\sin \left(\omega t + \frac{\pi}{2} \right).$$

The relations between the sinusoidal variations in e_c and i are described completely by the ratio of their amplitude values,

$$\left| \frac{\hat{E}}{\hat{I}} \right| = \frac{1}{\omega C} \quad \text{or} \quad \left| \frac{\hat{I}}{\hat{E}} \right| = \omega C \quad (5.4)$$

and the fact that the voltage sinusoid is one-quarter period ($\pi/2$ in phase angle) behind the current or the current is one-quarter period ahead of the voltage.

The ratios are unchanged if the current and voltage are measured in rms units. Both magnitude and phase relations are specified in the complex ratios (5.5), which give the impedance and admittance of the capacitor,

$$Z = \frac{E}{I} = \frac{1}{\omega C} / -90^\circ = -j \frac{1}{\omega C} \quad Y = \frac{I}{E} = \omega C / +90^\circ = j\omega C \quad (5.5)$$

where E and I denote the complex rms values of voltage and current and Z and Y the complex impedance and admittance.

6. Resistor.—Ohm's law applied to instantaneous values for an ideal resistor R gives

$$e_R = Ri \quad \text{or} \quad i = \frac{e_R}{R} \quad (6.1)$$

where e_R is the instantaneous voltage opposing the current. The shape of the current wave is the same as that of the voltage wave,

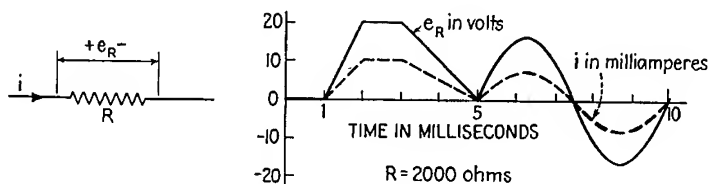


FIG. 6.1 —Ideal resistor and its instantaneous current-voltage relations.

Fig. 6.1b. The amplitudes of sinusoidal variations have the ratios

$$\frac{\hat{E}}{\hat{I}} = R \quad \text{or} \quad \frac{\hat{I}}{\hat{E}} = \frac{1}{R} \quad (6.2)$$

and the current and voltage are in phase. The impedance and admittance of a resistor are

$$Z = \frac{E}{I} = R / 0^\circ = R \quad \text{and} \quad Y = \frac{I}{E} = \frac{1}{R} / 0^\circ = \frac{1}{R} \quad (6.3)$$

7. Inductor.—A simple inductor consists of several turns of wire in the form of a coil. A changing current through such a coil, of self-inductance L , causes a potential difference in opposition to current change, Fig. 7.1, given by

$$e_L = L \frac{di}{dt} \quad (7.1)$$

the resistance of the coil being neglected. This relation is similar

to (5.1) for a capacitor, except that e and i are interchanged. This is seen also by comparison of Fig. 7.1b with Fig. 5.1b.

If the current variation is sinusoidal,

$$i = |\hat{I}| \sin \omega t \quad \text{and} \quad e_L = \omega L |\hat{I}| \sin \left(\omega t + \frac{\pi}{2} \right) \quad (7.2)$$

Comparison of the sinusoids of (7.2) shows that

$$\left| \frac{\hat{E}}{\hat{I}} \right| = \omega L \quad \text{or} \quad \left| \frac{\hat{I}}{\hat{E}} \right| = \frac{1}{\omega L} \quad (7.3)$$

and the voltage is one-quarter period ($\pi/2$ in phase angle) ahead of the current or the current is one-quarter period behind the

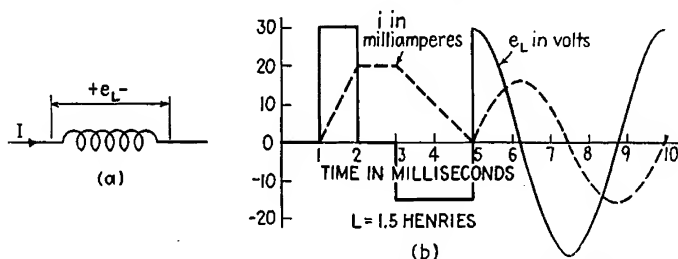


FIG. 7.1.—Ideal inductor and its instantaneous current-voltage relations.

voltage. The impedance and admittance for the inductor are

$$Z = \omega L / +90^\circ = j\omega L \quad \text{and} \quad Y = \frac{1}{\omega L} / -90^\circ = -j \frac{1}{\omega L} \quad (7.4)$$

III. SERIES CIRCUITS

8. Voltage and Current in a Series Combination.—In a series combination of ideal elements, Fig. 8.1, the instantaneous current through each element is the same. If the current is sinusoidal (has simple harmonic variation) the voltage across each element is sinusoidal.

The vector diagram above each element in Fig. 8.1 describes the magnitude and the phase of the sinusoidal voltage of that element with respect to the current. The complex descriptions of these voltage vectors in terms of the rms value I of the current are

$$E_c = -j \frac{1}{\omega C} I \quad E_R = RI \quad E_L = j\omega LI \quad (8.1)$$

The current-voltage ratios are the same whether they are measured in amplitude or in rms values. Unless instantaneous values

are needed, it is more convenient to use rms values, which will be used from now on unless otherwise stated.

The sinusoidal voltage across the series combination is represented by the vector sum of the voltages across the separate ele-

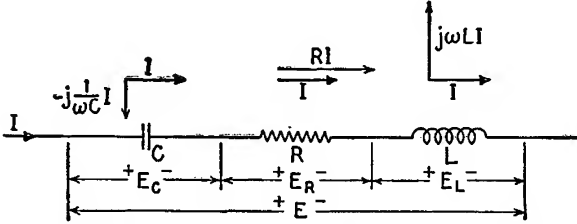


FIG. 8.1.—Voltages and their representative vectors in a series circuit.

ments. These voltages may be added graphically, Fig. 8.2a, or in terms of their complex representations,

$$E = -j\frac{1}{\omega C}I + RI + j\omega LI \quad (8.2)$$

As the voltages across the capacitor and inductor are oppositely

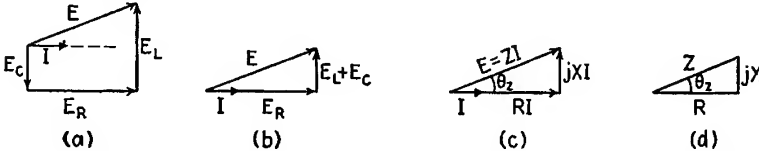


FIG. 8.2.—Current-voltage relations in a series circuit and their description in complex form.

directed (differ in phase by π radians, or 180°), they may be combined and the simplified Fig. 8.2b obtained, or

$$E = RI + j\left(\omega L - \frac{1}{\omega C}\right)I \quad (8.3)$$

From (8.3) the impedance of the combination can be expressed in terms of the circuit parameters as

$$Z = \frac{E}{I} = R + j\left(\omega L - \frac{1}{\omega C}\right) = R + jX \quad (8.4)$$

The impedance of the series combination is the *complex sum* of the impedances of the separate elements.

By definition the reactance X of the circuit is the coefficient of $+j$ in the complex expression for Z , or

$$E = (R + jX)I = ZI \quad (8.5)$$

The significance of Z is clarified if the expression is written in polar form

$$E = ZI = |Z|/\theta_z |I|/0^\circ = |Z| |I|/\theta_z$$

The magnitude $|E|$ of the voltage is $|Z|$ times the magnitude $|I|$ of the current through the circuit, and its phase with respect to the current is θ_z , Fig. 8.2*d*.

$$|Z| = \sqrt{R^2 + X^2} = \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2} \quad \text{and} \quad \tan \theta_z = \frac{X}{R} = \frac{\omega L - \frac{1}{\omega C}}{R} \quad (8.6)$$

The voltage across the impedance caused by the current through it is analogous to the RI drop in d-c circuits and may be referred to as the impedance drop. Extending this to (8.2) the voltage drop across the group is the *complex sum* of the impedance drops across its separate elements.

In (8.5) the product RI gives the voltage component in phase with I , while the product XI gives the voltage component whose phase differs from I by 90° . These components are referred to as the inphase and quadrature components of voltage. The power dissipated in the circuit is equal to the inphase component of voltage times the current, or $|RI^2|$. The complete circuit must contain an active element or source of electrical power to supply the energy

dissipated. The elements C , R , L , which are not sources of sustained power, are referred to as passive elements.

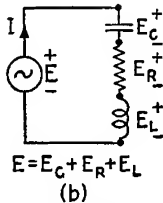
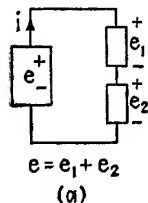


FIG. 9.1.—Kirchhoff's law applied to instantaneous and to complex voltages.

9. Kirchhoff's Voltage Law in an A-c Circuit.—Kirchhoff's voltage law holds for instantaneous values in an a-c circuit.

If e is the instantaneous terminal voltage of an electrical source, Fig. 9.1*a*, and e_1 and e_2 are the instantaneous voltages across passive elements, all measured by instruments whose polarities are as indicated, e is given by the algebraic sum of the meter readings,

$$e = e_1 + e_2 \quad (9.1)$$

If the three voltages are shown simultaneously on a cathode-ray

oscilloscope, the sum of e_1 and e_2 is equal to e at every instant of time, for any waveform of applied voltage and for any load.

If the applied voltage is sinusoidal and linear elements C , R , L constitute the load, (8.2) may be considered as describing the equality of the applied, or generator, terminal voltage and the sum of the sinusoidal voltages across the elements of the load. If these voltages be described as the complex products of the impedance of the element and the current,

$$E = Z_C I + Z_R I + Z_L I \quad (9.2)$$

where the subscripts denote the elements whose impedance is indicated. A circuit corresponding to (9.2) is shown in Fig. 9.1b. The generator voltage is assumed positive in the direction of the current, and the polarity of the voltages across the passive elements is in the direction opposite to the direction of the current. With these sign conventions,

$$E - Z_C I - Z_R I - Z_L I = 0 \quad (9.3)$$

which extends Kirchhoff's voltage law to the analysis of a-c circuits.

10. Series Circuits with Dissipative Reactors.—In actual capacitors and inductors dissipation of energy always accompanies

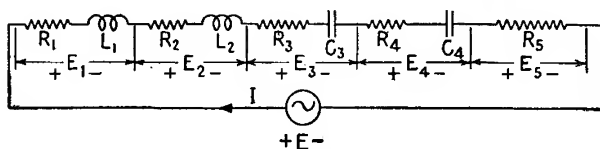


FIG. 10.1.—Series circuit with dissipative elements.

the passage of current. This loss of energy may be accounted for by associating a series resistance with each element. If several such elements constitute the load of a generator, Fig. 10.1, Kirchhoff's law gives

$$E = (R_1 + j\omega L_1)I + (R_2 + j\omega L_2)I + \left(R_3 - j\frac{1}{\omega C_3}\right)I + \left(R_4 - j\frac{1}{\omega C_4}\right)I + R_5 I \quad (10.1)$$

Collecting similar terms,

$$E = \left\{ (R_1 + R_2 + R_3 + R_4 + R_5) + j \left[(\omega L_1 + \omega L_2) - \left(\frac{1}{\omega C_3} + \frac{1}{\omega C_4} \right) \right] \right\} I \quad (10.2)$$

$$= (R + jX)I = ZI$$

The impedance of the series combination is the complex sum of the impedances of the separate elements.

11. Mutual Inductance in a Series Circuit.—When two coils are close together, the varying magnetic field of one coil may induce a voltage in the other coil. If the two coils of Fig. 11.1*a* are wound so that direct currents in the direction indicated in each produce magnetic flux in the same or aiding direction along their axes, the mutual inductance M is positive according to the convention of Sec. 13, Appendix B. If the coils are connected in series aiding, Fig. 11.1*b*, and an alternating current is sent through the combination, the induced voltage opposing the changing current in each coil is increased owing to the action of the other coil.

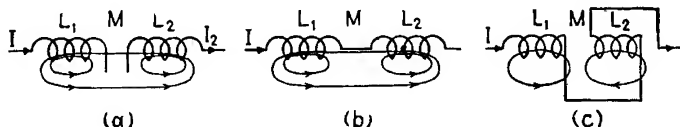


FIG. 11.1.—Schematic indication of the effects of mutual inductance.

If the connections of one coil are reversed, Fig. 11.1*c*, the magnetic fields along the axes are reduced. The value of M is negative. The equivalent self-inductance L of the combination is

$$L = L_1 + L_2 + 2M \quad (11.1)$$

The symbol M denotes a positive quantity for the aiding combination and a negative quantity for the opposing combination. The self-inductance of the reversed combination is less by $|4M|$ than that for the aiding combination.

IV. PARALLEL CIRCUITS

12. Voltage and Current in a Parallel Combination.—In a parallel circuit several paths, or branches, are provided for the current. The applied voltage across the group is identical in magnitude and phase with that across each branch. The current through each branch, Fig. 12.1, is determined wholly by the voltage across the branch and the impedance of the branch. If the generator voltage is constant, the current in each branch is constant. By an extension of Kirchhoff's current law to a-c circuits, the total current is the sum of the separate sinusoidal currents, or

$$I = I_C + I_R + I_L \quad (12.1)$$

$$= j\omega CE + \frac{1}{R} E - j \frac{1}{\omega L} E \quad (12.2)$$

$$\begin{aligned}
 &= \left[\frac{1}{R} - j \left(\frac{1}{\omega L} - \omega C \right) \right] E \\
 &= (G - jB)E = YE
 \end{aligned} \tag{12.3}$$

The complex number multiplying E is the admittance Y . Its real part is designated as G , the conductance of the combination, and the coefficient of $-j$ is defined as the susceptance B of the circuit.¹ The admittance of the combination is the sum of the admittances of the separate branches.

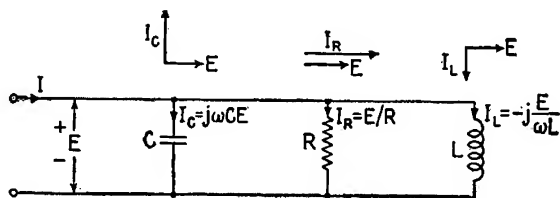


FIG. 12.1.—Parallel circuit with ideal elements.

Above each element, Fig. 12.1, is indicated the phase relation between the current through it and the common voltage vector. The vector sum of the branch currents is shown in Fig. 12.2. Each diagram of Fig. 12.2 represents the addition of the same currents; in Fig. 12.2c they are described in terms of the admittance, conductance, susceptance, and the applied voltage.

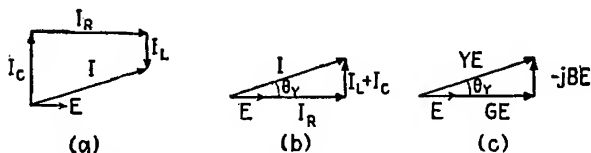


FIG. 12.2.—Vector sum of the branch currents.

13. Impedance of a Parallel Circuit.—If a circuit combination consists of several branches in parallel, the admittance of the group may be written as the sum of the admittances of the separate branches,

$$Y = Y_1 + Y_2 + Y_3 + \cdots \tag{13.1}$$

As admittance is the reciprocal of impedance, (13.1) may be written

$$\frac{1}{Z} = \frac{1}{Z_1} + \frac{1}{Z_2} + \frac{1}{Z_3} + \cdots \tag{13.2}$$

¹ American Standard Definitions of Electrical Terms, ASA C42-1941, American Institute of Electrical Engineers.

If the circuit has only two branches,

$$Z = \frac{Z_1 Z_2}{Z_1 + Z_2} = \left(\frac{\text{product}}{\text{sum}} \right) \quad (13.3)$$

The circuit of Fig. 13.1a consists of a parallel combination of an inductor, with an associated series resistance, and a capacitor

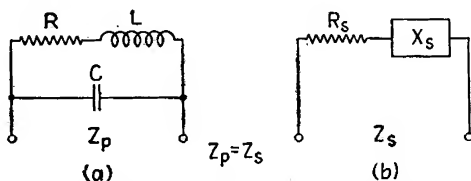


FIG. 13.1.—Two-branch parallel circuit and equivalent series circuit, whose elements are denoted by the subscript *s*.

if negligible dissipation. From (13.2),

$$\frac{1}{Z} = \frac{1}{R + j\omega L} + j\omega C \quad (13.4)$$

Reducing these to a common denominator,

$$\frac{1}{Z} = \frac{(1 - \omega^2 LC) + j\omega CR}{R + j\omega L} \quad (13.5)$$

Inverting this expression and rationalizing,

$$Z = \frac{R}{(1 - \omega^2 LC)^2 + R^2 C^2 \omega^2} + j \frac{\omega L(1 - \omega^2 LC - R^2 C/L)}{(1 - \omega^2 LC)^2 + R^2 C^2 \omega^2} \quad (13.6)$$

$$= R_s + jX_s \quad (13.7)$$

The real and imaginary parts of the expression for Z represent the resistance and reactance of a simple series circuit, Fig. 13.1b, equivalent to the parallel combination. The equivalent series resistance R_s and equivalent series reactance X_s are functions of the circuit constants and in general vary with the frequency.

14. Equivalent Representations of Dissipative Reactors.—A dissipative reactor can be represented as either a parallel or a series combination of reactance and resistance, Fig. 14.1. As the impedance Z_s of the equivalent series combination must be the same as the impedance Z_p of the equivalent parallel combination, the relations between R_s , X_s and R_p , X_p may be calculated from

$$R_s + jX_s = \frac{jR_p X_p}{R_p + jX_p} \quad (14.1)$$

Cross multiplying and equating reals and imaginaries,

$$R_s R_p = X_s X_p \quad (14.2)$$

$$X_s R_p + R_s X_p = R_p X_p \quad (14.3)$$

From these two independent conditions may be obtained a relation convenient for slide-rule computation,

$$X_s^2 + R_s^2 = |Z|^2 = R_s R_p = X_s X_p \quad (14.4)$$

The ratio of the series reactance to the series resistance of a reactor is defined as Q , its quality factor. The reciprocal of Q is

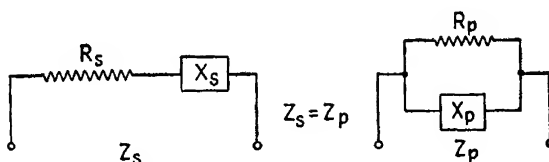


FIG. 14.1.—Alternative series and parallel representations of a practical circuit element.

its dissipation factor D . From (14.2),

$$Q = \frac{X_s}{R_s} = \frac{R_p}{X_p} = \frac{1}{D} \quad (14.5)$$

For specific elements, Q and D are

$$\begin{array}{ll} \text{for an inductor} & \text{for a capacitor} \\ Q = \frac{\omega L_s}{R_s} = \frac{R_p}{\omega L_p} = \frac{1}{D} & Q = \frac{1}{\omega C_s R_s} = \omega C_p R_p = \frac{1}{D} \end{array} \quad (14.6)$$

Note that both Q and D are explicit functions of frequency. If (14.3) be divided by $X_s R_p$ and $R_s X_p$, it follows from (14.5) that

$$1 + \frac{1}{Q^2} = \frac{X_p}{X_s} \quad \text{and} \quad 1 + Q^2 = \frac{R_p}{R_s} \quad (14.7)$$

If Q is larger than 10, the following approximations are accurate to within 1 per cent at least:

$$X_p \doteq X_s \quad R_p \doteq Q^2 R_s \quad (14.8)$$

Under these circumstances the relation between the equivalent series and parallel elements is very simple. The reactor has the same inductance or capacitance in either representation; a small series resistance is equivalent to a parallel resistance Q^2 times as large.

15. Current and Voltage Relations in a Parallel Circuit.—The currents through the two branches of the circuit, Fig. 15.1a, may be calculated separately. For the first branch,

$$I_1 = \frac{E}{Z_1} = \frac{E}{R_1 + j\omega L_1} = \frac{R_1 - j\omega L_1}{R_1^2 + \omega^2 L_1^2} E \quad (15.1)$$

This may be written as

$$I_1 = Y_1 E = \left(\frac{R_1}{|Z_1|^2} - j \frac{\omega L_1}{|Z_1|^2} \right) E = (G_1 - jB_1) E \quad (15.2)$$

where

$$G_1 = \frac{R_1}{|Z_1|^2} \quad \text{and} \quad B_1 = \frac{\omega L_1}{|Z_1|^2} \quad (15.3)$$

A similar analysis gives I_2 , the current in the second branch,

$$\begin{aligned} I_2 &= \frac{E}{Z_2} = \frac{E}{R_2 - j/\omega C_2} = \left(\frac{R_2}{|Z_2|^2} + j \frac{1/\omega C_2}{|Z_2|^2} \right) E \\ &= (G_2 - jB_2) E \end{aligned} \quad (15.4)$$

where

$$G_2 = \frac{R_2}{|Z_2|^2} \quad \text{and} \quad B_2 = \frac{-1/\omega C_2}{|Z_2|^2} \quad (15.5)$$

Each of the currents I_1 and I_2 consists of a sum of two components. The product GE gives the magnitude of the component

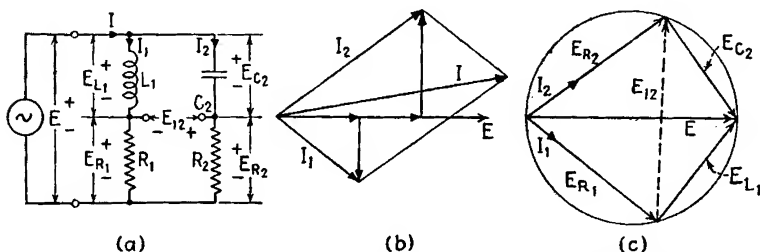


FIG. 15.1.—Two-branch parallel circuit and its current-voltage relations.

of current in phase with the voltage, and the product BE gives the component of current lagging the voltage by 90° . The power dissipated in either branch equals GE^2 .

The total current I is the sum of the complex currents,

$$I = I_1 + I_2 = [(G_1 + G_2) - j(B_1 + B_2)]E \quad (15.6)$$

The current relations are shown in Fig. 15.1b. The branch currents are added to give the total current. From the phase relations between currents and applied voltage, it is possible to construct

Fig. 15.1c, which describes the relations among the voltages across the elements in each branch. The voltages across the elements of each branch must have E as their vector sum. The voltage E_{R_1} across the resistor R_1 is in phase with I_1 , the current in R_1 . This allows E_{R_1} to be drawn parallel to I_1 . As the voltage E_{L_1} across L_1 leads E_{R_1} by 90° , they form a right triangle, with E as the hypotenuse. Therefore, their junction must lie on a circle with E as its diameter. A similar construction is shown for the voltages in branch 2, so that Fig. 15.1c is a complete vector diagram of the voltage-current relations in the circuit of Fig. 15.1a.

16. Electrical Representation of an A-c Generator.—An a-c generator acts as a linear element when its action upon a load can

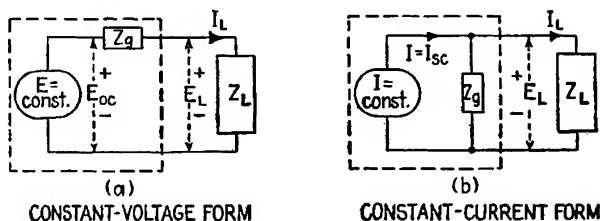


FIG. 16.1.—Alternative descriptions of a linear a-c generator.

be described in terms of a constant a-c voltage or current and a fixed internal impedance. These values may be dependent upon the frequency generated but are assumed to be independent of the load current.

If every value of load impedance Z_L applied to such a generator is very large in comparison with the internal impedance Z_g of the generator, the terminal voltage does not differ appreciably from its open-circuit voltage. Then the internal impedance Z_g may be neglected, and the generator may be approximately described by the single constant E_{oc} , its open-circuit voltage or emf. If the load impedance is very small in comparison with Z_g , the load current does not differ appreciably from the short-circuit current of the generator. Under such conditions the generator may be approximately described by the single constant I_{sc} , its short-circuit current, where $I_{sc} = E_{oc}/Z_g$. For all values of load impedance a linear generator may be described as a zero-impedance constant-voltage generator in series with an impedance Z_g , Fig. 16.1a, or as the combination of an infinite-impedance constant-current generator with the internal impedance Z_g in parallel, Fig. 16.1b. Z_g has the same value in these alternative representations of the same generator.

Calculations based on both circuits of Fig. 16.1 give the same values of current and voltage in the load.

$$I_L = \frac{E_{oc}}{Z_g + Z_L} \quad (16.1)$$

$$E_L = Z_L I_L = \frac{Z_L}{Z_g + Z_L} E_{oc} \quad \left. \begin{array}{l} \text{for the} \\ \text{constant-} \\ \text{voltage form,} \\ \text{Fig. 16.1a} \end{array} \right\} \quad (16.2)$$

$$E_L = \frac{Z_g Z_L}{Z_g + Z_L} I_{sc} \quad (16.3)$$

$$I_L = \frac{E_L}{Z_L} = \frac{Z_g}{Z_g + Z_L} I_{sc} \quad \left. \begin{array}{l} \text{for the} \\ \text{constant-} \\ \text{current form,} \\ \text{Fig. 16.1b} \end{array} \right\} \quad (16.4)$$

The expression for E_L in the constant-voltage form shows that the generator voltage E_{oc} divides between Z_g and Z_L in proportion to their complex impedances. In the constant-current form the expression for I_L shows that the generator current I_{sc} divides between Z_g and Z_L inversely proportional to their complex impedances. As $E_{oc} = Z_g I_{sc}$, the two expressions given for each of E_L and I_L are equivalent. The two representations of the generator differ in their predictions of power dissipated within the generator.

The description to be used is a matter of convenience and may be determined by the load characteristics. If the load consists only of series elements, the constant-voltage, or series, form of the generator simplifies the solution as Z_g represents merely another series element. If the load consists of parallel branches, the constant-current, or parallel, form makes Z_g merely an additional branch.

Similar representations for a d-c generator are explained in Appendix C.

CHAPTER II

CIRCUIT RESPONSE

The source of signal voltage in communication networks usually is greatly different, both physically and electrically, from the generator in power circuits. Its internal impedance may be so large in comparison with that of the load that it approximates a constant-current source. Its output in general is a combination of sinusoids whose frequencies, amplitudes, and phases vary with time. The communication circuit to which the signal is applied consists in part of simple units whose purpose is to act upon the input currents and voltages and deliver an altered output. The response characteristic of such a system describes the relationship between the amplitude and phase of the output and input for various frequencies or for different values of a variable circuit element.

I. RESPONSE OF SIMPLE LR AND CR UNITS

1. Variations of Z and Y for Circuits of Constant R .—When the generator approximates constant-current or constant-voltage operation over the useful range of the variable, the response of a unit is often simply related to its impedance or admittance. When L and R or C and R are in series and the only variable is the reactance, the impedance is

$$Z = R + jX = R \left(1 + j \frac{X}{R} \right) = R(1 + j \tan \theta_z) \quad (1.1)$$

The variations of Z are due only to changes in $\tan \theta_z$. Also,

$$|Z| = \frac{1}{|Y|} \quad \text{and} \quad \theta_z = -\theta_Y \quad (1.2)$$

The varying magnitudes of both Z and Y are determined by R and the value of $\tan \theta_z$;

$$|Z| = R \sqrt{1 + \tan^2 \theta_z} \quad \text{and} \quad |Y| = \frac{1}{R} \frac{1}{\sqrt{1 + \tan^2 \theta_z}} \quad (1.3)$$

The variations in magnitude and angle of Z and Y are indicated in Fig. 1.1, where for simplicity only positive values of X are shown.

As X increases from zero to infinity, Z always terminates upon a vertical line located at a distance R to the right of the origin, Fig. 1.1a. This line is a locus describing all possible values of Z . The corresponding values of Y terminate upon a semicircular locus¹ of diameter $1/R$ as shown in Fig. 1.1c.

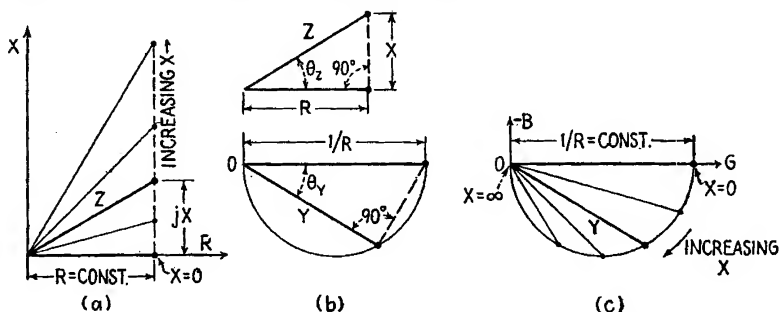


FIG. 1.1.—Impedance and admittance relations in a series circuit with constant R : (a) values of Z as X is varied; (b) relation between Z and its corresponding Y ; (c) values of Y corresponding to indicated values of Z .

2. Variations of Z and Y for Simple LR and CR Units.—The impedance and admittance of a series or parallel LR or CR combination may be varied by changing the frequency. The expressions for Z and Y are simplified if the impressed angular frequency $\omega = 2\pi f$ is compared with the angular frequency ω' that makes the reactance equal in magnitude to the resistance. This frequency ω' is related under some circumstances to the power dissipation and is referred to as the half-power frequency of the LR or CR combination.

For an LR combination

$$\omega' = \frac{1}{L/R}$$

For a CR combination

$$\omega' = \frac{1}{CR} \quad (2.1)$$

In Chap. VI on Transients, the quantities L/R and CR are shown to be the time constants of LR and CR circuits. The impedance of the series combination and the admittance of the parallel combination for the four possible arrangements of R in series or parallel with L or C are as follows:

¹ The triangle associated with Y in Fig. 1.1b is similar to the triangle in the diagram for Z , since the sides of the triangle for Y are proportional to the hypotenuse and the base of the triangle for Z , and the included angles are equal. It is a right triangle, and for all values of Y corresponding to positive values of X it is inscribed in the semicircle in Fig. 1.1c. For negative values of X , the triangle of which Y is the hypotenuse is inscribed in a semicircle of the same diameter but above the OG axis of Fig. 1.1c.



(a)

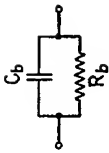
$$Z_a = R_a + j\omega L_a$$

$$= R_a \left(1 + \frac{j\omega L_a}{R_a} \right)$$

$$= R_a \left(1 + j \frac{\omega}{\omega'} \right)$$

$$Y_a = \frac{1}{R_a \left(1 + j \frac{\omega}{\omega'} \right)}$$

$$\tan \theta_z = \frac{\omega}{\omega'}$$



(b)

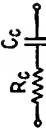
$$Y_b = \frac{1}{R_b} + j\omega C_b$$

$$= \frac{1}{R_b} \left(1 + j\omega C R_b \right)$$

$$= \frac{1}{R_b} \left(1 + j \frac{\omega}{\omega'} \right)$$

$$Z_b = \frac{R_b}{\left(1 + j \frac{\omega}{\omega'} \right)}$$

$$\tan \theta_y = \frac{\omega}{\omega'}$$



(c)

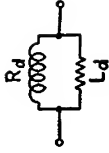
$$Z_c = R_c - j \frac{1}{\omega C_c}$$

$$= R_c \left(1 - j \frac{1}{\omega C R_c} \right)$$

$$= R_c \left(1 - j \frac{\omega'}{\omega} \right)$$

$$Y_c = \frac{1}{R_c \left(1 - j \frac{\omega'}{\omega} \right)}$$

$$\tan \theta_z = - \frac{1}{\omega/\omega'}$$



(d)

$$Y_d = \frac{1}{R_d} - j \frac{1}{\omega L_d}$$

$$= \frac{1}{R_d} \left(1 - j \frac{R_d}{\omega L_d} \right)$$

$$= \frac{1}{R_d} \left(1 - j \frac{\omega'}{\omega} \right)$$

$$Z_d = \frac{R_d}{\left(1 - j \frac{\omega'}{\omega} \right)}$$

$$\tan \theta_y = - \frac{1}{\omega/\omega'}$$

The frequency ratio ω/ω' determines $\tan \theta_z$ or $\tan \theta_Y$ and therefore θ_z and θ_Y . The corresponding values of $|Z|$ and $|Y|$ are given by (1.3) in terms of the constant R and either $\sqrt{1 + \tan^2 \theta_z}$ or its equal $\sqrt{1 + \tan^2 \theta_Y}$.

The expressions (2.3) are not affected by interchanging ω with L or C . Then ω/ω' can be replaced in (2.4), (2.5), and (2.6) by L/L' if L is varied or by C/C' if C is varied. Here L' or C' may be defined as the value of L or C at which the reactance equals the resistance in magnitude, or as the half-power value of L or C , equal to R/ω or $1/\omega R$, respectively.

The expressions in (2.4a) and (2.4b) for the series LR and the parallel CR circuits show that their variations are identical provided that Y be exchanged for Z and $1/R$ for R . The variations of $|Z_a|$ and $|Y_b|$ for the series LR and the parallel CR circuits above are both represented by a single curve, shown at the upper left of Fig. 2.1; the numerical scale represents the ratio of the variable to its half-power value. The corresponding variations of $|Y_a|$ and $|Z_b|$ are shown in the lower left of the figure. The polar descriptions are completed by giving the corresponding values of θ_z and θ_Y . The rectangular components of Z and Y are shown in addition. At the right of Fig. 2.1, curves are given for the series CR and parallel LR circuits (c) and (d).

When these LR and CR combinations serve as loads for constant-current or constant-voltage generators, these curves may also be employed to show the current and voltage response and the variations in power. The vertical scales are changed in accordance with the following equations to yield the quantities named, the horizontal ratio scale remaining unchanged:

Constant-voltage generator		Constant-current generator	
Load current	$= EY$	Load voltage	$= IZ$
Power dissipated	$= E^2G$	Power dissipated	$= I^2R$
Reactive power	$= E^2B$	Reactive power	$= I^2X$
Phase of I with respect to E	$= \theta_Y$	Phase of E with respect to I	$= \theta_z$

The variations in the quantities represented by the curves of Fig. 2.1 are summarized by the linear or circular locus at the left of each set of curves. Values of Z and Y are shown identified with the corresponding numerical values of ω/ω' , or L/L' , or C/C' , depending upon the variable. The scale of each locus is determined by the value of R , and the angle for any impedance Z is

given by the corresponding value of $\tan \theta_z$ from (2.6). The impedance of a series combination or the admittance of a parallel combination has a linear locus. Its reciprocal has a semicircular locus, on the opposite side of the axis of reals. For increasing values of any one of ω , L , or C , the numerical values of the corresponding ratios ω/ω' , L/L' or C/C' increase upward along the linear loci and clockwise around the semicircular loci.

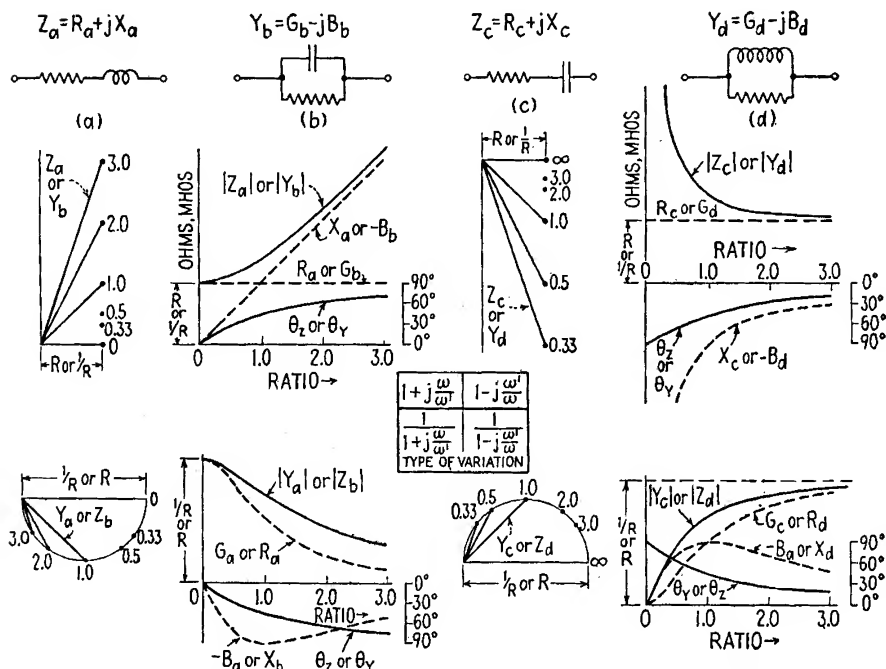


FIG. 2.1.—Impedance and admittance variations of LR and CR combinations as ω , L , or C is varied. Numerical ratio scale represents ω/ω' , L/L' , or C/C' .

3. Variation of Load Voltage with Constant-current Generator.

The voltage output of a device is of primary importance when applied to a vacuum tube. Under many conditions of operation the tube has a high input impedance, and its behavior is determined by the applied voltage.

When the LR and CR combinations of Fig. 3.1 serve as loads for a constant-current generator, the voltage E developed across them is equal to ZI . The changes in magnitude and phase of E are due only to changes in Z . The ordinates of the curves for $|Z|$ in Fig. 2.1 multiplied by $|I|$ yield the corresponding variations

of $|E|$ shown in Fig. 3.1. The value of θ_z describes the phase of the load voltage with respect to the current. In the series LR circuit, $|E|$ rises continuously as the values of the variable become large, and in the series CR circuit $|E|$ rises as the values of the variable become small owing to the corresponding increase in $|Z|$.

It should be remembered that for any actual generator there is a limit to the value of $|Z|$ above which constant-current operation cannot be approximated.

If the complex values of Z shown in the loci of Fig. 2.1 be multiplied by the constant current I , the resulting vector voltages are identified by the same numerical values of the ratios. All

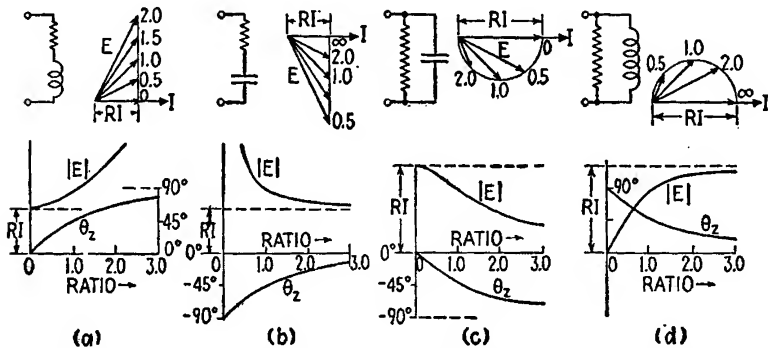


FIG. 3.1.—Voltage variation across LR and CR loads driven by a constant-current generator. R is constant. Numerical scale is ratio ω/ω' , L/L' , or C/C' , according to which is the variable.

possible values of load voltage therefore have loci of the same shape and ratio scale as those of Z .

4. Voltage-divider Response.—If a generator maintains a constant input voltage E_i across a series LR or CR combination, the voltage across each series element varies with ω , L , or C , and these circuits may be used as “voltage dividers” to yield an output voltage adjustable in magnitude and phase. The four types of such simple voltage dividers are illustrated in Fig. 4.1.

The output voltage E_o across R in the CR series circuit is

$$E_o = \frac{R}{R - j\frac{1}{\omega C}} E_i = Y R E_i = \frac{1}{1 - j\frac{\omega'}{\omega}} E_i \quad (4.1)$$

where the admittance Y of the series CR circuit is the only variable, and from (4.1) and (2.5c) the output voltage response across

R has the form of Y_c in Fig. 2.1. This curve is reproduced with suitable change in legend in Fig. 4.1a.

The voltage across R in the LR series circuit is equal to a constant times the admittance of the series LR combination. This admittance is Y_a in Fig. 2.1, and the output voltage E_o across R has the same variation as in Fig. 4.1b. The curves are identical except for a change of the vertical scale.

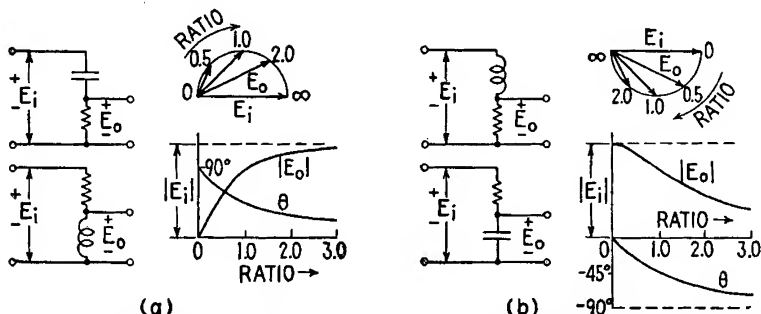


FIG. 4.1—Output voltage variation of voltage dividers. R is constant. Numerical scale is ratio ω/ω' , L/L' , or C/C' , according to which is the variable.

The output voltage across L in the LR series circuit is

$$E_o = \frac{j\omega L}{R + j\omega L} E_i = \frac{1}{1 - j \frac{R}{\omega L}} E_i = \frac{1}{1 - j \frac{\omega'}{\omega}} E_i \quad (4.2)$$

and has the same variations (4.1) as the voltage across the resistor in a CR voltage-divider circuit. Similarly, the voltage variation across C in the RC voltage divider is the same as that across R in the LR voltage-divider circuit. Two curves therefore suffice to describe the response of the four voltage-divider circuits.

The two loci in Fig. 4.1 describe the vector relations between the input and output voltage for different numerical values of the ratios ω/ω' , L/L' , or C/C' , whichever is the variable.

5. Generator of Intermediate Resistance.—In many cases of importance the impedance of the generator is a pure resistance R_g of the same order of magnitude as the load impedance. Because of its magnitude, R_g then must be taken into account in either the constant-current or the constant-voltage representation of the generator. If the load consists of elements in parallel, as in Fig. 5.1b, the constant-current representation of the generator simplifies the problem, as the internal resistance R_g is then represented in parallel with the load resistance R and the two may be combined to form a single parallel resistance R_p . The voltage across the load

due to the generator current $I = E_{oc}/R_g$ is equal to

$$E_L = \frac{-jR_p \frac{1}{\omega C}}{R_p - j \frac{1}{\omega C}} I = \frac{R_p}{1 + j\omega CR_p} I \quad (5.1)$$

Since

$$R_p = \frac{RR_g}{R + R_g} \quad \text{and} \quad \omega' = \frac{1}{CR_p} = \frac{1}{C \left(\frac{RR_g}{R + R_g} \right)} \quad (5.2)$$

then

$$E_L = \frac{RR_g}{R + R_g} \frac{1}{\left(1 + j \frac{\omega}{\omega'} \right)} \frac{E_{oc}}{R_g} = \frac{R}{R + R_g} \frac{1}{\left(1 + j \frac{\omega}{\omega'} \right)} E_{oc} \quad (5.3)$$

The frequency response is similar to that of Fig. 3.1c, where constant current was assumed through the *load*. The only difference is the reduction of the output voltage in the ratio $R/(R + R_g)$ and the altered value of ω' as given by (5.2).

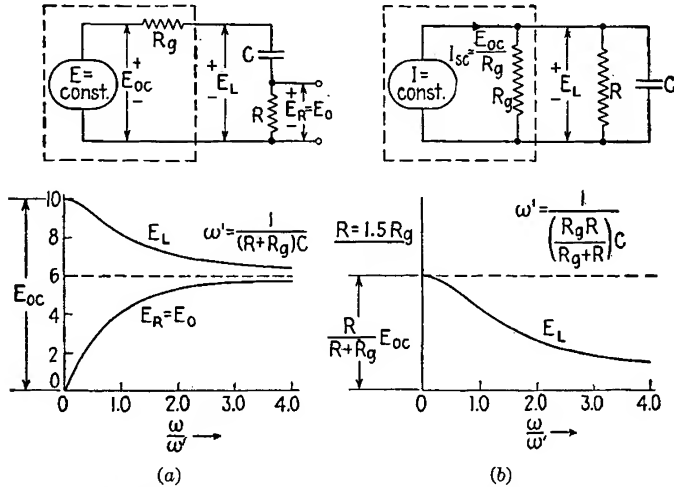


FIG. 5.1.—A generator represented in series form with a series load, and in parallel form with a parallel load. Voltage curves assume the same generator for each case, with $R_g = \frac{1}{3}R$.

If the load is a series combination, Fig. 5.1a, the constant-voltage representation of the generator allows R_g to be added to the series resistance R , giving an equivalent series resistance R_s . The voltage response across R is similar to that obtained with no generator resistance. Only part of the voltage across R_s appears

across R , however, and ω' equals $1/CR_s$. The curve E_R of Fig. 5.1a and the curve E_L of Fig. 5.1b have the same shape as the low-frequency and high-frequency response of a resistance-capacitance coupled amplifier, Chap. XIII.

The frequency-response curve for the load voltage $|E_L|$ across R and C , Fig. 5.1a, differs from the corresponding curve $|E|$, Fig. 3.1b. At zero frequency it approaches not an infinite but a finite value, equal to E_{oc} , and its slope can be adjusted by choice of R_g/R to yield a more slowly changing response characteristic.

6. Compensated Voltage Divider.—For many purposes it is necessary to decrease an applied voltage in a ratio that is independent of frequency, preferably without change in phase angle.

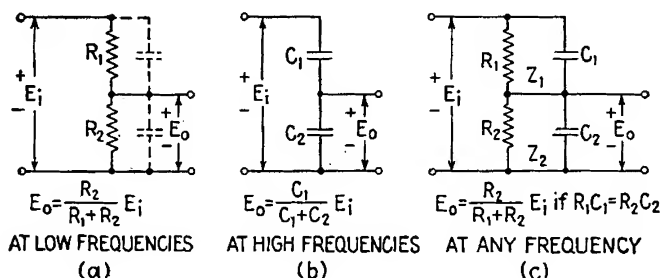


FIG. 6.1.—Voltage-divider circuits.

This can be done at low frequencies by the resistance voltage divider of Fig. 6.1a. As the frequency is increased, however, the impedance of the stray capacitance across each series element decreases, causing the attenuation to vary with frequency. This can be avoided at high frequencies by the capacitive voltage divider of Fig. 6.1b. Stray capacitance across each series capacitor simply increases its capacitance, and the capacitive admittance at high frequencies is chosen large enough to make any shunting conductance negligible.

If parallel capacitors are added to the resistive voltage divider, the circuit of Fig. 6.1c is obtained, where C_1 and C_2 include the stray capacitances across R_1 and R_2 . If the capacitors are adjusted so that the half-power frequencies of the two resistor-capacitor combinations are the same,

$$C_1 R_1 = C_2 R_2 \quad \text{and} \quad \frac{E_o}{E_i} = \frac{Z_2}{Z_1 + Z_2} = \frac{R_2}{R_1 + R_2} \quad (6.1)$$

and the voltage division is constant at all frequencies. Here

Z_1 and Z_2 are the impedances of the parallel RC units. By making R_1 and R_2 large and keeping C_1 and C_2 as small as possible, high input impedance as well as constant voltage division may be maintained over a wide range of frequencies.

7. Phase-shifting Networks.—In the previous LR and CR circuits, changes in the phase angle of I or E were accompanied by a change in magnitude. By the use of special circuits it is possible to change the phase of a sinusoid without appreciable change in magnitude, although in many cases the voltage is reduced by a constant factor.

A center-tapped resistor connected in parallel with a series CR combination forms the phase-shifting circuit of Fig. 7.1. The voltages E_R across R and E_C across C when added form a right

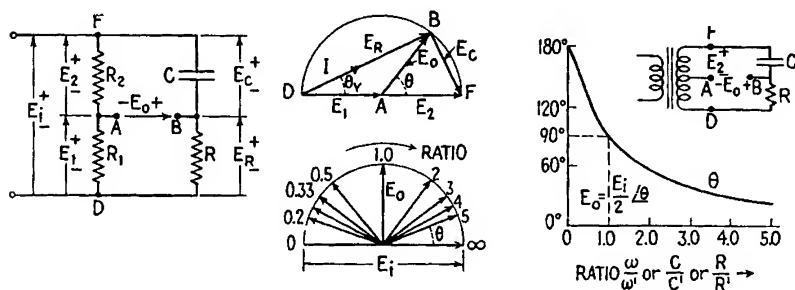


FIG. 7.1.—Phase-shifting network and its voltage relationships.

triangle with E_i as the hypotenuse. This triangle is inscribed within a semicircle with E_i as its diameter. The voltages across the equal resistances R_1 and R_2 are in phase with E_i , and their sum is equal to E_i . If the center tap of the pure resistance branch is A and the junction of R and C is B , a vector drawn from A to B represents the output voltage E_o of the phase shifter. When the impedance of R or C is varied, the relative magnitudes of E_R and E_C are altered, but their vector diagram is still a triangle inscribed in the same semicircle. The tip of E_o moves around the circumference of the semicircle when R , ω , or C is varied. The numbers on the lower semicircle of Fig. 7.1 are the numerical ratios of any one of R , ω , or C (whichever is the variable) to its half-power value. The constant magnitude of the output voltage is $E_i/2$. The phase angle of the output voltage with respect to the input voltage can be varied by this circuit from an angle near 0° for large values of the variable to an angle of 180° for zero value of the variable.

The combination of the source E_i and the center-tapped resistive

branch can be replaced by the center-tapped secondary of a transformer, resulting in the circuit at the right of Fig. 7.1. If the center tap is grounded, E_{AB} and E_2 share a common ground and are equal in magnitude. The phase of E_{AB} may be altered in the same manner as in the circuit with the center-tapped resistor.

II. SERIES RESONANCE

It is generally possible to adjust the circuit constants or the impressed frequency in series circuits containing L , C , and R so that the phase angle between current and voltage is zero and the impedance of the combination is a pure resistance. This condition is defined as resonance. At resonance the voltage across the circuit is in phase with the current through it, and the power factor of the circuit is unity.

8. Circuit Relations at Series Resonance.—In the series circuit of Fig. 8.1 the internal impedance of the generator is assumed negligible in comparison with R , so that R accounts for all the circuit dissipation. The current is

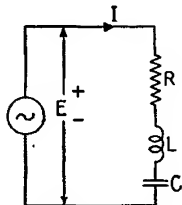


FIG. 8.1.—Series circuit.

$$I = YE = \frac{E}{Z} = \frac{E}{R + j\left(\omega L - \frac{1}{\omega C}\right)} \quad (8.1)$$

At resonance the current and voltage are in phase, and the coefficient of j in (8.1) therefore must be zero. The condition for resonance is

$$\left(\omega L - \frac{1}{\omega C}\right) = 0 \quad \omega L = \frac{1}{\omega C} \quad \text{or} \quad \omega^2 LC = 1 \quad (8.2)$$

This condition may be satisfied by variation of any one of the parameters involved in (8.2). If these are altered separately, there are certain critical values of ω , L , or C that produce resonance. These values, denoted by the subscript r , are related to the fixed constants of the circuit by

$$\omega_r = \frac{1}{\sqrt{LC}} \quad L_r = \frac{1}{\omega^2 C} \quad \text{and} \quad C_r = \frac{1}{\omega^2 L} \quad (8.3)$$

The series-resonant frequency f_r of the circuit, in cycles/second, is related to ω_r by $\omega_r = 2\pi f_r$.

At resonance, the impedance Z and current I reduce to

$$Z = R \quad \text{and} \quad I = \frac{E}{R} \quad (8.4)$$

The voltages across R , L , and C are

$$E_R = RI = E \quad E_L = j \frac{\omega L}{R} E \quad \text{and} \quad E_C = -j \frac{E}{\omega CR} \quad (8.5)$$

From (8.2) the voltages E_L and E_C at resonance are equal in magnitude and opposite in phase.

The equal voltage ratios $|E_L/E|$ and $|E_C/E|$ at resonance, obtained from (8.5), are given by the ratio of the magnitude of the reactance of either the inductor or the capacitor to the series resistance of the circuit. The definition of Q for a dissipative reactor (the ratio of its series reactance to series resistance) can be extended to a series circuit containing L , R , and C . The quality factor Q_r of a series branch or loop is defined as the ratio of the series reactance of either reactor *at resonance* to the series resistance of the branch or loop. If the frequency is variable,

$$Q_r = \frac{\omega_r L}{R} = \frac{1}{\omega_r CR} = \frac{1}{R} \sqrt{\frac{L}{C}} \quad (8.6)$$

where ω_r is the angular frequency at resonance.

For a circuit of constant L , R , and C , the quality factor Q_r is also a constant, independent of frequency. If the resistance R is primarily that of the inductor, Q_r for the circuit is equal to the quality factor Q of the coil at the resonant frequency of the circuit. At resonance, from (8.5),

$$|E_L| = |E_C| = Q_r |E|$$

When Q_r is large, E_L and E_C are many times greater than the applied voltage.

9. Energy and Power Relations at Series Resonance.—An inductor carrying current and a charged capacitor possess stored energy. The instantaneous value of the stored energy in each reactor is

$$\begin{array}{ll} \text{in the inductor} & \text{in the capacitor} \\ \frac{1}{2} Li^2 & \frac{1}{2} C e_c^2 \end{array} \quad (9.1)$$

As the sinusoidal current through each reactor in a series circuit is the same and since the voltage across the capacitor lags the current

by one-quarter period, the stored energy in the capacitor is a maximum one-quarter cycle later than that in the inductor. From (8.2) and (8.5) the maximum energy stored by the inductor and capacitor is the same at resonance, as indicated in Fig. 9.1a. The sum of the stored energy is a constant, equal to the maximum energy stored in either L or C ; its value is $L|\hat{I}|^2/2$ in amplitude units or $L|I|^2$ in rms units. The ratio of the stored energy at resonance to the energy dissipated per cycle is $Q_r/2\pi$.

The transfer of energy from inductor to capacitor and back occurs twice per period of the current. The rate of transfer, or the instantaneous power absorbed from and returned to the circuit by

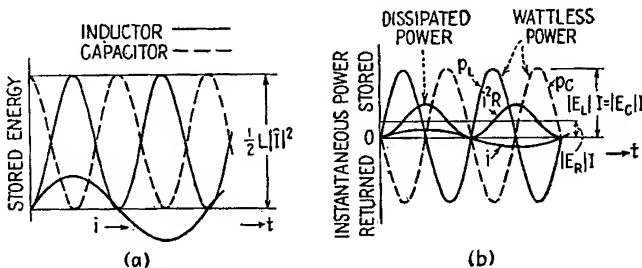


FIG. 9.1.—Stored energy and power variations at resonance.

each reactor, is sketched in Fig. 9.1b, together with curves describing the instantaneous and average power dissipation in the resistor. At resonance, the instantaneous rate of energy storage by one reactor is equal at all times to the rate of energy return by the other reactor. The maximum rate of energy storage in both reactors may be calculated from (4.3), Chap. I, as $|\hat{E}_C \hat{I}|/2 = |\hat{E}_L \hat{I}|/2$, or in rms values $|E_C I|$ and $|E_L I|$. As the currents and voltages are in quadrature, these expressions give the reactive volt-amperes, or so-called “wattless power,” for the capacitor and inductor.

The ratio of the reactive volt-amperes for either reactor to the power dissipated at resonance is equal to Q_r . This fact may also be used to define the quality factor Q_r .

10. The Frequency Response of a Series LCR Circuit.—As Z is equal to the voltage across a circuit per unit current through it, the voltage variation with constant current is determined by the variation of Z . The impedance Z of a series LCR circuit is given by

$$Z = R + jX = R + j\left(\omega L - \frac{1}{\omega C}\right) \quad (10.1)$$

and

$$|Z| = \sqrt{R^2 + X^2} \quad \theta_z = \tan^{-1} \frac{\omega L - \frac{1}{\omega C}}{R} \quad (10.2)$$

The magnitudes of the components of Z are plotted in Fig. 10.1 as a function of frequency. At low frequencies the capacitive reactance is predominant, and at high frequencies the inductive reactance is predominant. At some intermediate frequency these reactances cancel; the capacitive and inductive reactances are equal and

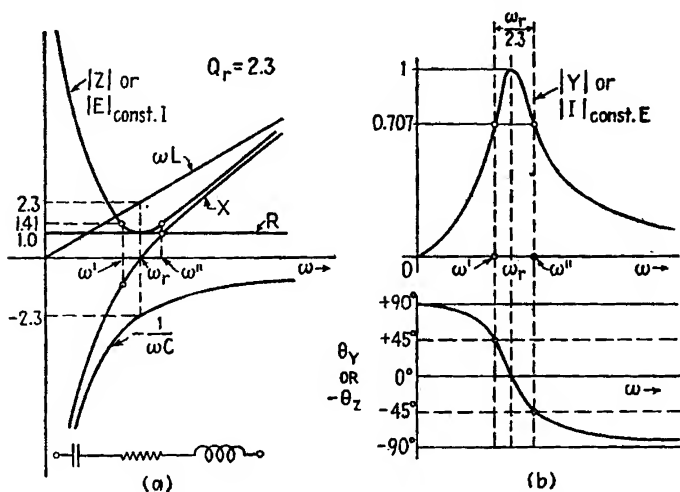


FIG. 10.1.—Variations in Z , Y , and θ_Y for a series LCR circuit as the frequency is changed.

opposite in sign. The frequency at which this occurs is the resonant frequency f_r ; the reactance X is zero, and the impedance Z is a minimum and is equal to R . The rapidity with which $|Z|$ changes in the neighborhood of resonance depends upon Q_r , which is the ratio at resonance of either the capacitive or the inductive reactance to the resistance. The rapidity of this change is measured by the separation of the lower and upper half-power angular frequencies ω' and ω'' marked on the curves. At these frequencies X equals $-R$ and R , respectively, and the magnitude of the impedance is $\sqrt{2}$ times its resonant value R .

As the admittance of a circuit is equal to the current per unit voltage applied, the curves for $|Y|$ and θ_Y in Fig. 10.1b describe the variations in magnitude and phase of the current in a series LCR circuit when a constant-voltage generator of variable frequency is

applied to it. The values of Y may be obtained from those for Z by the relations $|Y| = 1/|Z|$ and $\theta_Y = -\theta_Z$. The admittance is a maximum at the resonant frequency and drops to $1/\sqrt{2}$ of this value at the half-power frequencies, at which the power delivered to the circuit is one-half that at resonance. The phase angle θ_Y at ω' and ω'' is $+45^\circ$ and -45° , respectively.

The effect of changing the Q_r of the circuit is illustrated by plotting several response curves for different values of the circuit

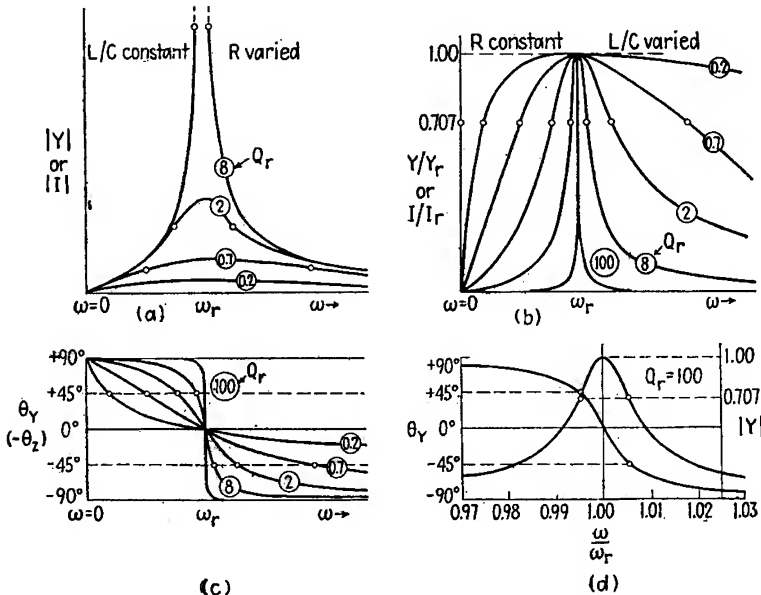


FIG. 10.2.—Response curves of a series LCR circuit as a function of frequency for various values of Q_r (encircled).

constants. In Fig. 10.2a the L/C ratio is fixed, and R is changed from curve to curve. At frequencies removed from resonance, the reactance is predominant, and therefore the reactance controls the shape of the curve; the dependence on R is slight. Near resonance, the resistance controls the shape of the curve in a region that is smaller the lower the value of R or the higher the value of Q_r . In Fig. 10.2b the resistance is held constant and the L/C ratio changed. The larger the value of L/C and therefore the larger the value of Q_r , the more restricted the range of large response. The sharpness of tuning, which varies inversely with the separation of the half-power frequencies, is correspondingly increased.

Most of the values of Q_r shown in Fig. 10.2 are small. It would be difficult to illustrate in Fig. 10.2a the response for a circuit with a Q_r as large as the readily obtained value of 100, as the response would not differ appreciably from that of $Q_r = 8$ for the lower values of $|Y|$ or $|I|$, which are the only values shown on the diagram. In Figs. 10.2b,c the response for $Q_r = 100$ is indicated, but the changes in $|Y|$ and θ_Y occur in such a narrow range of frequencies that the shape is obscured. In Fig. 10.2d a portion of the response curve with $Q_r = 100$ is drawn to an expanded frequency scale, which covers a frequency variation of a few per cent in the neighborhood of resonance. The curves for both $|Y|$ and θ_Y are nearly symmetrical about ω_r in this range.

11. Calculation of Z and Y for a Range of Frequencies.—The impedance of a series LCR circuit may be written

$$Z = R + j\left(\omega L - \frac{1}{\omega C}\right) = R \left[1 + j\left(\frac{\omega L}{R} - \frac{1}{\omega CR}\right)\right] \quad (11.1)$$

From this expression the angle and magnitude of Z are

$$\tan \theta_Z = \frac{\omega L}{R} - \frac{1}{\omega CR} \quad \text{and} \quad |Z| = R \sqrt{1 + \tan^2 \theta_Z} \quad (11.2)$$

The corresponding admittance Y is

$$|Y| = \frac{1}{R} \frac{1}{\sqrt{1 + \tan^2 \theta_Z}} \quad \text{and} \quad \theta_Y = -\theta_Z \quad (11.3)$$

If R is constant and ω is varied, the changing magnitudes of Z and Y are determined by the value of $\tan \theta_Z$. In terms of ω_r and Q_r , the resonant frequency and quality factor of the circuit, the expression for $\tan \theta_Z$ is

$$\tan \theta_Z = \left(\frac{\omega_r L}{R} \frac{\omega}{\omega_r} - \frac{1}{\omega_r CR} \frac{\omega_r}{\omega}\right) = Q_r \left(\frac{\omega}{\omega_r} - \frac{\omega_r}{\omega}\right) \quad (11.4)$$

where

$$Q_r = \frac{\omega_r L}{R} = \frac{1}{\omega_r CR} = \frac{1}{R} \sqrt{\frac{L}{C}} \quad (11.5)$$

From (11.4) it follows that the variations of $\tan \theta_Z$, and therefore those of Z and Y , are determined by the circuit constant Q_r and by the ratio of the impressed angular frequency ω to the series-resonant angular frequency ω_r . If the frequency response of a series circuit is plotted in terms of this frequency ratio, all circuits with

the same Q_r yield curves of the same shape. As the frequency ω occurs only as compared with ω_r , it is convenient to refer to this ratio as w and call it the fractional frequency. Then

$$\tan \theta_z = Q_r \left(w - \frac{1}{w} \right) \quad \text{where} \quad w = \frac{\omega}{\omega_r} = \omega \sqrt{LC} \quad (11.6)$$

The relation (11.6) is useful if it is desired to compare the response at some frequency removed from resonance with that obtained at resonance. For example, if a circuit is tuned to a signal delivered by a constant-voltage generator, the current response at a frequency twice as great (the second harmonic of the signal, Chap. IX) is readily calculated from

$$w = 2 \quad \tan \theta_z = Q_r(2 - \frac{1}{2}) = 1.5Q_r \quad |Z| = R \sqrt{1 + (1.5Q_r)^2} \quad (11.7)$$

The value of $|Y|$, and therefore the current, is reduced to

$$\frac{1}{\sqrt{1 + 2.25Q_r^2}}$$

of its value at resonance, and it lags the voltage by $\tan^{-1} 1.5Q_r$. If $1.5Q_r$ is large, $|Y|$ is reduced by the approximate factor $1/(1.5Q_r)$, or $1/\tan \theta_z$. If a circuit having the same Q_r were tuned to the second harmonic of a signal, its response at the fundamental frequency ($w = \frac{1}{2}$) would be reduced by the same factor, since for a given value of w and its reciprocal the magnitude of $\tan \theta_z$ in (11.6) is the same.

The half-power frequencies of the circuit may be determined in terms of Q_r by solving (11.6) for w , yielding the following quadratic expression and its solution:

$$w^2 - \frac{\tan \theta_z}{Q_r} w - 1 = 0 \quad w = \frac{\tan \theta_z}{2Q_r} + \sqrt{1 + \left(\frac{\tan \theta_z}{2Q_r} \right)^2} \quad (11.8)$$

Since at resonance $w = 1$ and $\tan \theta_z = 0$, only the positive sign of the radical has meaning. At the half-power frequencies the following relations hold:

Upper half-power frequency	Lower half-power frequency
$\omega = \omega'' \quad w = \frac{\omega''}{\omega_r} \quad \tan \theta_z = 1$	$\omega = \omega' \quad w = \frac{\omega'}{\omega_r} \quad \tan \theta_z = -1$
$\frac{\omega''}{\omega_r} = \frac{1}{2Q_r} + \sqrt{1 + \frac{1}{4Q_r^2}}$	$\frac{\omega'}{\omega_r} = \frac{-1}{2Q_r} + \sqrt{1 + \frac{1}{4Q_r^2}}$

(11.9)

Upon subtracting these expressions, the radical vanishes and

$$\frac{\omega'' - \omega'}{\omega_r} = \frac{1}{Q_r} \quad \text{or} \quad Q_r = \frac{\omega_r}{\omega'' - \omega'} = \frac{\omega_r}{BW} \quad (11.10)$$

where BW is the difference between the half-power frequencies and is called the band width. It may be measured in radians/second, equal to $\omega'' - \omega'$, or cycles/second, equal to $f'' - f'$. The ratio of the band width to the resonant frequency is called the fractional band width and is $BW/\omega_r = 1/Q_r$. The band width in per cent of the resonant frequency is $100/Q_r$. The expressions (11.10) are exact and hold for any value of Q_r .

In some circuits the applied signal voltage contains many frequencies closely grouped around the resonant frequency of the circuit. When the per cent deviation from the resonant frequency is small, an approximate relation between the frequency and $\tan \theta_z$ simplifies the calculation of the circuit response.

If $(\tan \theta_z)/2Q_r$ is small in comparison with unity, its square in the radical of (11.8) may be neglected and

$$w - 1 = \frac{\omega - \omega_r}{\omega_r} \doteq \frac{\tan \theta_z}{2Q_r} \quad (11.11)$$

It follows that $(\tan \theta_z)/2Q_r$ approximates the ratio of the frequency deviation from resonance to the resonant frequency when this ratio is small. Under these conditions the frequency deviation from resonance is

$$\omega - \omega_r \doteq \frac{1}{2} \frac{\omega_r}{Q_r} \tan \theta_z = \frac{BW}{2} \tan \theta_z \quad (11.12)$$

At the half-power frequencies $\tan \theta_z = \pm 1$, and the approximate values of ω'' and ω' are

$$\omega'' \doteq \omega_r + \frac{BW}{2} \quad \text{and} \quad \omega' \doteq \omega_r - \frac{BW}{2} \quad (11.13)$$

They occur at approximately a half band width above and below the resonant frequency. At intervals of 2, 3, 4 half band widths above and below the resonant frequency, $\tan \theta_z$ has the value ± 2 , ± 3 , ± 4 ; from (11.3) the corresponding values of admittance are $1/\sqrt{5}$, $1/\sqrt{10}$, $1/\sqrt{17}$ of its value at resonance. Where this approximation is applicable, $\tan \theta_z$ is readily obtained as the ratio of the frequency deviation from resonance to the half band width.

$$\tan \theta_z \doteq \frac{\omega - \omega_r}{BW/2} \quad (11.14)$$

This approximation predicts a symmetrical variation of $|Y|$ around the resonant frequency, as illustrated in Fig. 11.1*a* for a circuit of $Q_r = 50$. In the small frequency range of 4 per cent below and above resonance shown, the true curve has a negligible difference from the symmetrical one. Expansion of the radical in (11.8) shows that at a frequency near resonance the error involved in the use of the approximation of (11.12) is about one-half the per cent deviation from resonance.

For larger per cent deviation from resonance an accurate representation of $|Y|$ requires that points on the symmetrical curve be

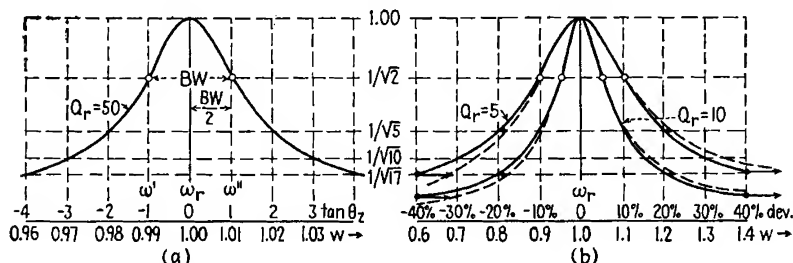


FIG. 11.1.—Variation of the admittance of a series LCR circuit near resonance, showing the departure from symmetry as the per cent deviation from resonance increases.

shifted toward higher frequency by an amount which depends only upon the per cent deviation from resonance and is independent of Q_r . In Fig. 11*b* the symmetrical and true curves are shown for circuits of Q_r equal to 5 and 10. At equal per cent deviations either above or below resonance the frequency difference between the two curves is the same and increases linearly with the deviation from resonance.

The simultaneous variations of magnitude and phase of Z and Y are illustrated by the loci of Fig. 11.2. Figure 11.2*a* shows the values of Z and $\tan \theta_z$ for a fixed R , at various frequencies corresponding to integral values of $\tan \theta_z$. Near resonance these frequencies have approximately equal differences of one-half band width when Q_r is large, *i.e.*, when the band width is a few per cent of the resonant frequency. In Fig. 11.2*b*, values of Z are related to w for a circuit of $Q_r = 5$ (20 per cent band width). For the large percentage frequency deviations indicated, the corresponding variations in Z are unsymmetrical. In Fig. 11.2*c* are shown the

values of Y corresponding to Figs. 11.2*a*, *b*. The symmetrical variation of Y near resonance in the upper circular locus is characteristic of high- Q_r circuits, while the unsymmetrical behavior for $Q_r = 5$ is characteristic of low- Q_r circuits.

If the LCR circuit is excited by a constant-current generator, the various values of Z in Fig. 11.2 multiplied by the current magnitude describe the variation of the vector voltage across the circuit both in magnitude and in phase with respect to the current.

When the circuit is excited by a constant-voltage device, the values of Y describe the current variations. Although the current

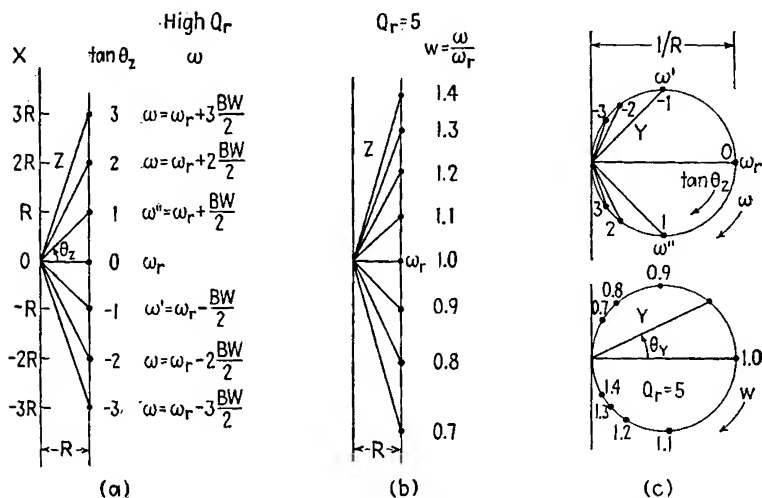


FIG. 11.2.—Variation of Z and Y as related to $\tan \theta_z$ and ω for a high- Q_r series LCR circuit and to w for a low- Q_r series circuit.

in a series LCR circuit is a maximum at resonance, the voltages across C and L are, respectively, a maximum at a frequency slightly below and slightly above resonance owing to the fact that their impedance varies with the frequency. For large values of Q_r the frequency deviation from resonance for these maxima is one part in $4Q_r^2$, and the maxima are larger than $Q_r E$ (the voltage at resonance) by only one part in $8Q_r^2$.

12. Variation of Capacitance in a Series LCR Circuit.—If R , L , and ω are fixed in a series LCR circuit and C is varied, the corresponding changes in Z and Y are similar to those caused by varying ω . The chief difference is in the variation of $\tan \theta_z$. As the frequency is constant and C is varied, the quality factor Q_r , defined at resonance, must be given in terms of C_r . The expression for

$\tan \theta_z$ from (11.2) becomes

$$\tan \theta_z = Q_r \left(1 - \frac{C_r}{C} \right) \quad \text{where} \quad Q_r = \frac{1}{R} \sqrt{\frac{L}{C_r}} = \frac{1}{\omega C_r R} = \frac{\omega L}{R} \quad (12.1)$$

For large values of C , $\tan \theta_z$ approximates Q_r , and the corresponding value of Y approaches $1/\sqrt{1 + Q_r^2}$ of its resonant value. From (12.1) the ratio C_r/C is

$$\frac{C_r}{C} = 1 - \frac{\tan \theta_z}{Q_r} \quad \text{and} \quad \frac{C}{C_r} = \frac{1}{1 - \frac{\tan \theta_z}{Q_r}} \quad (12.2)$$

At the upper and lower half-power values of capacitance, $\tan \theta_z$ equals $+1$ and -1 , respectively,

$$C'' = \frac{C_r}{1 - \frac{1}{Q_r}} \doteq C_r + \frac{C_r}{Q_r} \quad \text{and} \quad C' = \frac{C_r}{1 + \frac{1}{Q_r}} \doteq C_r - \frac{C_r}{Q_r} \quad (12.3)$$

The approximation shows that the half-power values are approximately equally spaced on either side of C_r if Q_r is large. The approximate per cent deviation of the half-power values of C from C_r is $100/Q_r$, or twice the percentage deviation of the half-power values of ω from ω_r , Sec. 11.

It is possible to determine Q_r exactly in terms of C'' and C' for any value of Q_r . At the half-power values of capacitance, (12.2) becomes

$$\frac{C_r}{C''} = 1 + \frac{1}{Q_r} \quad \text{and} \quad \frac{C_r}{C'} = 1 - \frac{1}{Q_r} \quad (12.4)$$

$$\frac{C_r}{C''} - \frac{C_r}{C'} = \frac{2}{Q_r} \quad \text{and} \quad \frac{C_r}{C'} + \frac{C_r}{C''} = 2 \quad (12.5)$$

$$\frac{C_r(C'' - C')}{C'C''} = \frac{2}{Q_r} \quad \text{and} \quad \frac{C_r(C'' + C')}{C'C''} = 2 \quad (12.6)$$

Dividing the second equation by the first in (12.6),

$$\frac{C'' + C'}{C'' - C'} = Q_r \quad (12.7)$$

Now $Q_r = \omega L/R$, (12.1), and $C_r = 1/\omega^2 L$. If these values are substituted in (12.6)

$$R = \frac{C'' - C'}{2\omega C'C''} \quad \text{and} \quad L = \frac{C'' + C'}{2\omega^2 C'C''} \quad (12.8)$$

Knowledge of the half-power capacitances permits the calculation by means of (12.8) of R and L at the frequency of operation.

III. PARALLEL RESONANCE

Parallel resonance is said to exist when the impedance of a parallel combination of reactive elements is a pure resistance, with zero phase angle. When driven by a generator, the reactive components of the branch *currents* add to zero, as contrasted to series resonance where the reactive components of *voltage* cancel.

Despite this and other differences there exist marked similarities in the behavior of the same elements excited in series and parallel resonance. Many of the series relationships are useful in describing conditions in the neighborhood of parallel resonance, particularly if the losses of the circuit are low.

13. Conditions at Parallel Resonance.—In the parallel combination of Fig. 13.1*a*, driven by a constant-voltage generator, it is

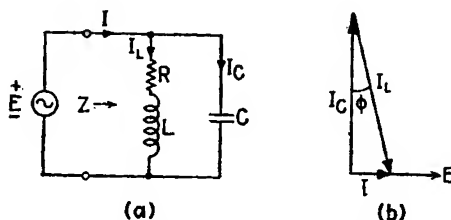


FIG. 13.1.—Parallel LCR circuit and its current-voltage relations at resonance.

assumed that the losses in the capacitor are negligible in comparison with those in the inductor. For suitable values of L , C , R , some value of ω produces parallel resonance. If R is small, the parallel-resonant frequency is practically the same as the series-resonant frequency of the loop; at the parallel-resonant frequency, the reactances of the two branches are nearly equal, and the branch currents, which differ in phase by almost 180° , tend to cancel at the generator terminals. If R were reduced to zero, the generator current would reduce to zero, the combination would have infinite resistance, and there would be a circulating current in the series loop equal to that through either of its branches. With finite though small values of R , the conditions at resonance differ somewhat from this ideal case, but the magnitudes of the branch and loop currents are not greatly affected.

The impedance of the parallel combination may be expressed as the product of the branch impedances divided by their sum,

$$Z = \frac{\left(-j \frac{1}{\omega C}\right)(R + j\omega L)}{R + j\left(\omega L - \frac{1}{\omega C}\right)} \quad (13.1)$$

$$= \frac{\frac{L}{C} \left(1 - j \frac{R}{\omega L}\right)}{R + j\left(\omega L - \frac{1}{\omega C}\right)} \quad (13.2)$$

$$= \frac{L}{CR} \frac{1 - j \frac{R}{\omega L}}{1 + j\left(\frac{\omega L}{R} - \frac{1}{\omega CR}\right)} \quad (13.3)$$

At resonance the impedance is a pure resistance, and the angle of the complex ratio multiplying L/CR in (13.3) is zero. This requires that the angles of the complex numerator and denominator be equal, then the tangents of these angles are equal, and

$$-\frac{R}{\omega L} = \frac{\omega L}{R} - \frac{1}{\omega CR} \quad (13.4)$$

Multiplying through by $\omega R/L$ and solving for ω yields ω_{pr} , the parallel-resonant angular frequency,

$$\omega_{pr} = \sqrt{\frac{1}{LC} - \frac{R^2}{L^2}} \quad (13.5)$$

The ratio of the parallel-resonant frequency to the series-resonant frequency of the loop ($\omega_{sr} = 1/\sqrt{LC}$) is

$$\frac{\omega_{pr}}{\omega_{sr}} = \sqrt{1 - \frac{R^2 C}{L}} = \sqrt{1 - \frac{1}{Q_r^2}} \doteq 1 - \frac{1}{2Q_r^2} + \dots \quad (13.6)$$

where Q_r is the quality factor of the series loop. The approximation given holds only for high- Q_r circuits where the resonant frequencies are almost the same. For $Q_r = 10$, the approximation indicates the parallel-resonant frequency is only $\frac{1}{2}$ per cent lower than the series-resonant frequency.

The parallel-resonant frequency is a function of R . In fact, if R is so large that $Q_r = 1$, the parallel-resonant frequency is zero, from (13.6). For larger values of R no parallel-resonant frequency exists for this circuit,¹ and the ratio of driving frequency

¹ There is no frequency at which the total alternating current I , Fig. 13.1a, is in phase with the applied voltage E .

to parallel-resonant frequency has no meaning. The behavior of this parallel LCR circuit is described, therefore, in terms of the previously defined Q_r and the frequency ratio $w = \omega/\omega_{sr}$ of the series loop. Although at parallel resonance the ratio of reactive volt-amperes to power dissipated is not equal to Q_r , it does not differ sensibly from it for a high- Q_r circuit.

The parallel-resonant impedance Z_{pr} is obtained by substituting (13.4) in (13.3). The impedance of the parallel combination at the series-resonant frequency of the loop, $(Z)_{\omega_{sr}}$, is obtained from (13.1) by setting $\omega = 1/\sqrt{LC}$. The impedances presented to the generator at these two important frequencies are

$$Z_{pr} = \frac{L}{CR} \quad Z_{\omega_{sr}} = \frac{L}{CR} - j\sqrt{\frac{L}{C}} \quad (13.7)$$

Multiplication by R/R yields from (8.6) the same relations expressed in terms of Q_r ,

$$\left. \begin{aligned} Z_{pr} &= Q_r^2 R \\ Z_{\omega_{sr}} &= Q_r^2 R - jQ_r R \\ &= Q_r^2 R \left(1 - j\frac{1}{Q_r} \right) \end{aligned} \right] \quad (13.8)$$

Where parallel resonance exists, the parallel-resonant impedance and the real part of the parallel impedance at ω_{sr} are equal. The phase angle θ_z for the combination at ω_{sr} is always negative since $\tan \theta_z = -1/Q_r$ at this frequency.

From (13.7) the smaller the value of R , the larger the parallel-resonant impedance L/CR at the parallel-resonant frequency. This is because the currents in the two branches at ω_{pr} become more nearly equal in magnitude and opposite in phase as R is decreased. The relation at parallel resonance between the generator and branch currents and the generator voltage is shown in Fig. 13.1b. At parallel resonance the magnitude of the current through the inductor must be larger than that through the capacitor because its out-of-phase component is equal to that through the capacitor and, in addition, it possesses a component in phase with the generator voltage. As Q_r increases, the currents I_L and I_C approach equality of magnitude and opposition of phase. The generator current at resonance is thus progressively decreased, and the impedance of the combination becomes larger and larger. The sine of the angle ϕ of Fig. 13.1 is equal to $1/Q_r$, and at parallel resonance the current through the inductor is equal to Q_r times

the generator current for any value of Q_r . For large values of Q_r , the magnitude of the current through the capacitor approaches the magnitude of the current through the inductor, and the following relations obtain:

$$Q_r |I| = |I_L| \doteq |I_C| \quad (13.9)$$

Thus at parallel resonance there is a "resonant rise" of current, comparable with the "resonant rise" of voltage at series resonance. For a parallel-resonant circuit whose Q_r is high the circulating current in the LCR loop approximates Q_r times the generator current.

Q_r is as defined in (8.6). It is the Q of the "loop" at the *series-resonant* frequency of the loop.

14. Response of a High- Q Parallel Circuit near Resonance.—

If the parallel circuit of Fig. 13.1 is excited by a constant-current variable-frequency generator, the variation in voltage across the LCR combination is due to its changing impedance. The expression (13.2) for the impedance of the parallel circuit may be rearranged to give

$$Z = \left(\frac{L}{C}\right) \left[\frac{1}{R + j\left(\omega L - \frac{1}{\omega C}\right)} \right] \left(1 - j\frac{R}{\omega L}\right) \quad (14.1)$$

The quantity in the brackets on the right is the admittance Y_s of the series loop formed by the branches, so that

$$Z = \left(\frac{L}{C}\right) (Y_s) \left(1 - j\frac{1}{wQ_r}\right) = |Z|/\theta_z \quad (14.2)$$

where $w = \omega/\omega_{sr} = \omega\sqrt{LC}$ and Q_r is the quality factor of the series loop at series resonance. If $wQ_r > 1$, Q_r large and frequency deviation from resonance small, the last term in (14.1) is practically equal to unity and can be neglected.

Therefore, except for a constant multiplier L/C , the variation of $|Z|$ for the parallel circuit is practically the same as that of $|Y_s|$ for the series circuit. The symmetrical curves of $|Z|$ and θ_z , Fig. 14.1b, for a parallel circuit of $Q_r = 100$, do not differ appreciably in shape, for small deviations from resonance, from the curves for $|Y|$ and θ_y for a series circuit of the same Q_r , Fig. 10.2d. The band width and half-power frequencies of the series loop calculated from (11.10) and (11.13) also describe the variations of Z for the parallel combination.

From the values of $|Z|$ and θ_z for the parallel combination the

equivalent series resistance R_s and the equivalent series reactance X_s are obtained as the real and imaginary parts of Z . The complex rectangular expressions for R_s and X_s are given by (13.6), Chap. I, and R_s and X_s are shown as dashed lines in Fig. 14.1b. The equivalent series resistance R_s , which accounts for the power delivered to the load in terms of the total current I , Fig. 13.1a, varies more rapidly than $|Z|$ and falls to one-half its maximum value at the half-power frequencies. At frequencies below resonance the circuit is inductive, while at frequencies above resonance the circuit is capacitive, as shown by the negative value of X_s . The maximum values of X_s occur at the half-power frequencies.

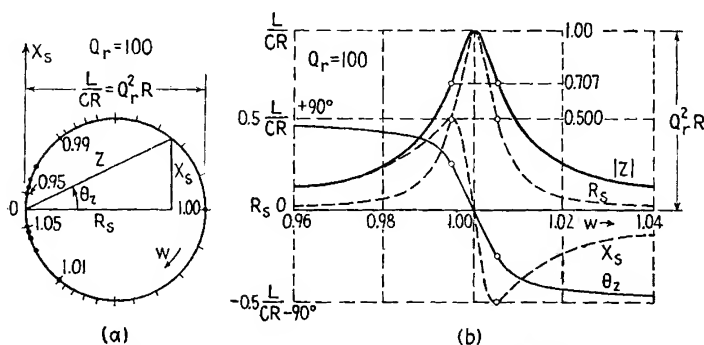


FIG. 14.1.—Impedance variations for a parallel circuit of high Q_r described in polar and rectangular form as a function of the frequency ratio $w = \omega/\omega_{sr} = \omega \sqrt{LC}$.

The variations illustrated in the four curves of Fig. 14.1b are described by the single circular locus of Z in Fig. 14.1a. Here a numerical scale relates the terminal positions of Z on its circular locus to w , the ratio of the exciting frequency to the series-resonant frequency of the loop.

When the circuit is connected to a constant-current generator, the voltage is given by the product of Z , Fig. 14.1, and the constant current, so that both the diagrams and the locus describe the variation in voltage across a high- Q , LCR combination as the frequency is altered. Near resonance the impedances of the LR and C branches change by only a few per cent, and the magnitudes of the current through both branches, or the loop current, follow the voltage variation closely so that the loop current near resonance is approximately proportional to $|Z|$, Fig. 14.1.

15. Response of Intermediate and Low- Q Parallel Circuits.—With low- Q , circuits and wide frequency deviations from resonance

the complex numeric factor of (14.2) cannot be neglected. If it is expressed in polar form,

$$1 - j \frac{1}{wQ_r} = \sqrt{1 + \frac{1}{(wQ_r)^2}} \angle \tan^{-1} \left(-\frac{1}{wQ_r} \right) \quad (15.1)$$

its action is seen to increase $|Z|$ by a factor dependent upon $(wQ_r)^2$ and to decrease θ_z by an angle determined by wQ_r . For small values of $1/(wQ_r)$ the change in angle is more important than the change in magnitude. This is illustrated in Fig. 15.1b, where the variation of $|Z|$ and θ_z is given for a circuit of $Q_r = 8$. The shape of the curve for Z differs only slightly from the curve for Y_s , while

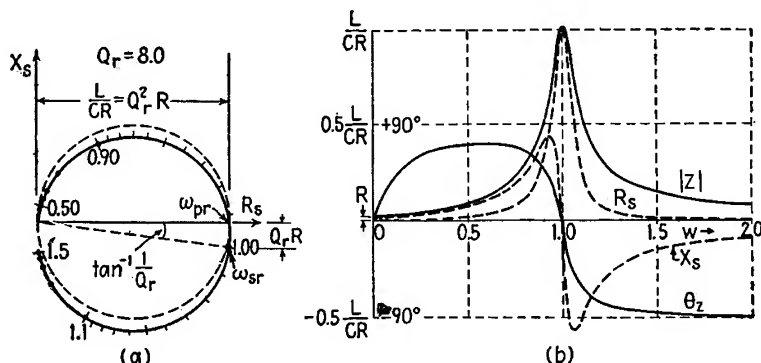


FIG. 15.1.—Impedance variations for a parallel circuit of intermediate Q_r described in polar and rectangular form as a function of the frequency ratio $w = \omega/\omega_{sr} = \omega \sqrt{LC}$.

the curve for θ_z is markedly affected. Therefore $|Z|$ falls to 0.707 of its maximum value at frequencies only slightly different from the half-power frequencies of the series loop; the ratio of the resonant frequency to the band width is equal to Q_r to a good approximation. The phase angle, however, no longer has equal magnitudes at the upper and lower half-power frequencies of the loop. Nor has X_s . These changes appear in the locus of Z , Fig. 15.1a, which, remaining almost circular, is slightly enlarged and rotated clockwise through an angle whose tangent approximates $1/Q_r$. At ω_{sr} the tangent of the phase angle is exactly $1/Q_r$, and the locus shows that ω_{sr} and ω_{pr} differ slightly in frequency in accordance with (13.6). The effect of the complex numeric factor is a maximum at zero frequency ($w = 0$) where $|Z|$ equals R and θ_z equals zero, Fig. 15.1b.

For still smaller values of Q_r the curve for $|Z|$ deviates sharply

from that of $|Y_s|$. In Fig. 15.2b are shown the variations for a circuit of $Q_r = 2$. The general behavior of $|Z|$ is the same as in Fig. 15.1b except near zero frequency, but θ_z is zero at a frequency well below ω_{sr} , and the broadened peaks of R_s and $|Z|$ have separate maxima at frequencies different from ω_{pr} . The maximum value of R_s , which corresponds to maximum delivery of power by a constant-current generator, occurs about halfway between ω_{pr} and ω_{sr} . The frequency giving a maximum value of $|Z|$ is only slightly below the series-resonant frequency and differs from it by approximately one part in $4Q_r^4$.

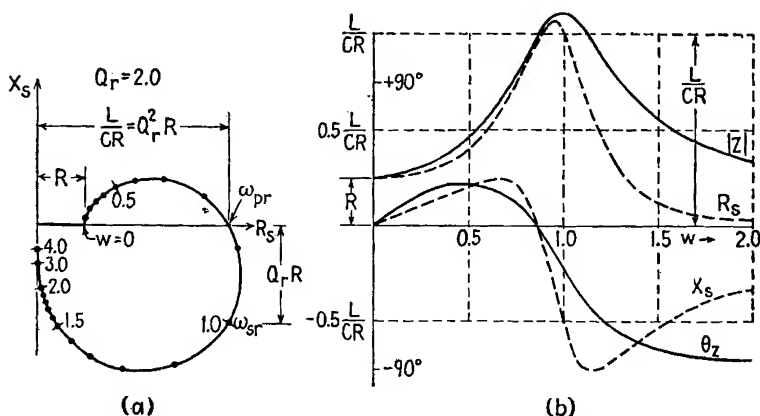


FIG. 15.2.—Impedance variations of a parallel circuit of low Q_r .

The behavior of the low- Q_r circuit shown in Fig. 15.2 exhibits many of the characteristics of the parallel combination that are not observable for larger values of Q_r .

16. Fractional Values of Q_r .—With decreasing values of Q_r the parallel-resonant frequency approaches zero, and the circuit is inductive over a decreasing range of frequencies near zero. At a Q_r of unity, Fig. 16.1a, both θ_z and X_s have negative values only and remain near zero over a considerable range of frequency. Both R_s and $|Z|$ still exhibit a resonant rise at a frequency somewhat below series resonance for the loop.

With further reduction of Q_r to 0.707, Fig. 16.1b the maximum of R_s vanishes, and R_s decreases slowly from its zero-frequency value of R . The curve for $|Z|$ rises slightly in a broad flat maximum and returns to the value R at 0.707 of the series-resonant frequency of the loop.¹ The phase angle θ_z decreases steadily and varies

¹ This corresponds to $N = 0.5$ in Chap. XIII, Sec. 18.

almost linearly with frequency although a trace remains of the inflection shown in the $Q_r = 1$ curve. For a Q_r of 0.5, Fig. 16.1c, the decrease of θ_z is more rapid and almost linear. The curvature at the origin has reversed, indicating that some intermediate value

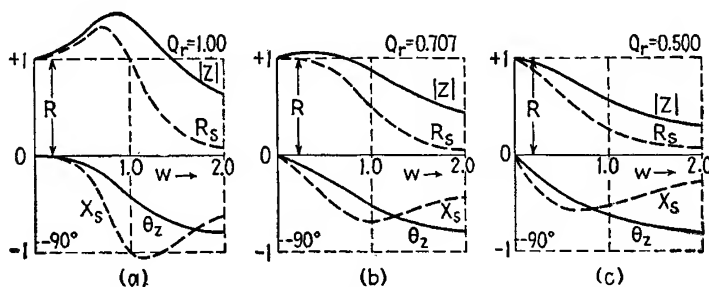


FIG. 16.1.—Impedance variations for a parallel circuit of fractional Q_r .

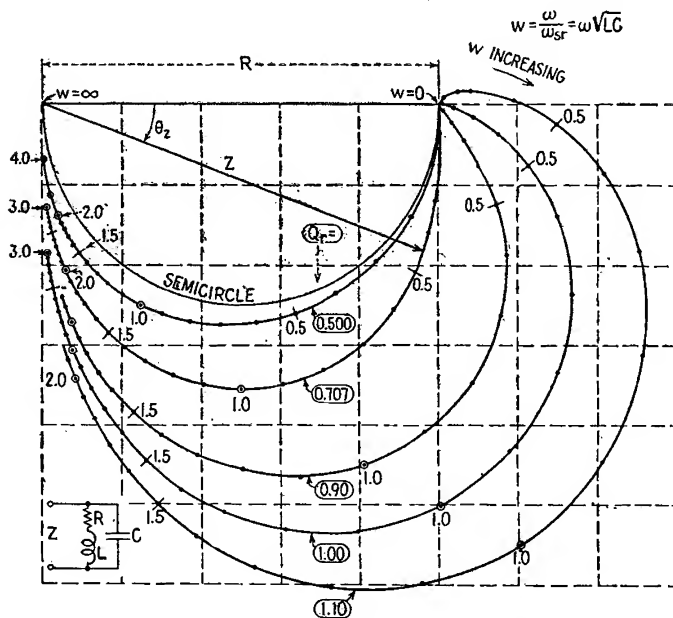


FIG. 16.2.—Impedance variations for parallel circuits having very low values of Q_r . of Q_r (approx. 0.57) would ensure linear variation of the phase angle over a wide range of frequencies beginning at zero.¹

The progressive changes in the variations of Z are also illustrated by the loci of Fig. 16.2 drawn for values of Q_r slightly above unity, unity, and ranging down to zero. The semicircular locus

¹ This corresponds to $N = 0.32$ in Chap. XIII, Sec. 18.

for $Q_r = 0$ is that of a capacitor and resistor in parallel, as L is reduced to zero. The loci also show that $Q_r = 1$ ensures that the circuit simulates a pure resistance with nearly zero phase angle over a range of frequencies near zero; the magnitude of its impedance is changing, however. The 0.707 locus gives an almost constant value of $|Z|$ although θ_z changes rapidly near zero frequency.

17. Division of Resistance between Branches in a Parallel Circuit.—If the resistance R is transferred from the inductive to the capacitive branch, the curves previously drawn serve to describe the behavior of the new circuit for all values of Q_r , with the following modifications: The numbers on the frequency-ratio scale in terms

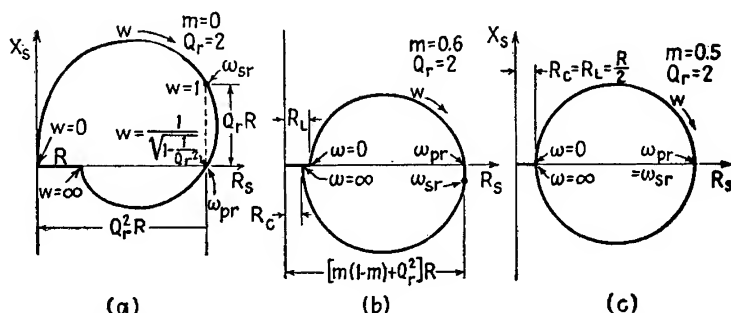


FIG. 17.1.—Impedance variations for a parallel circuit (a) with all the resistance in the capacitive branch, $m = 0$; (b) with unequal division of resistance, $m = 0.6$; (c) with equal division of resistance, $m = 0.5$.

of $w = \omega/\omega_{sr}$ are replaced by their reciprocals ω_{sr}/ω , that is, ∞ becomes $1/\infty = 0$, 5 becomes $\frac{1}{5}$, etc. The impedance at the parallel- and series-resonant frequencies is unchanged in magnitude; but the ratio of ω_{pr} to ω_{sr} (13.6) is inverted, and ω_{pr} is now larger than ω_{sr} , which is unchanged. The signs of θ_z and X_s are reversed. In terms of the loci this is equivalent to inverting them about the axis of reals and replacing w by its reciprocal. Compare Fig. 17.1a with Fig. 15.2a.

If resistance exists in both branches, the locus becomes a compromise between the two loci of Figs. 17.1a and 15.2a, as shown in Fig. 17.1b. The shape depends upon the division of resistance between the two branches. If $R = R_L + R_C$ is the total loop resistance and m is the fraction of the total resistance in the inductive branch, the characteristics of the circuit at parallel resonance may be described in terms of m and Q_r , the quality factor of the series loop. The ratio of parallel- to series-resonant frequency is

$$\frac{\omega_{pr}}{\omega_{sr}} = \sqrt{\frac{Q_r^2 - m^2}{Q_r^2 - (1 - m)^2}} \doteq 1 - \frac{2m - 1}{2Q_r^2} \quad (17.1)$$

The impedance at the parallel-resonant frequency is

$$Z_{pr} = \frac{R_L R_C}{R} + \frac{L}{CR} = [m(1 - m) + Q_r^2]R \doteq \frac{L}{CR} = Q_r^2 R \quad (17.2)$$

The approximations indicated hold for large values of Q_r .

If the resistance is equally divided between the branches ($m = 0.5$), the parallel- and series-resonant frequencies are the same and the locus reduces to a circle, Fig. 17.1c. At this resonant frequency both $|Z|$ and R_s have equal maximum values. If the resistance remains equally divided and the Q_r of the series loop is reduced by varying the L/C ratio, the diameter of the circular locus is reduced. At $Q_r = 0.5$, corresponding to $R_L = R_C = \sqrt{L/C}$, the locus reduces to a point and the circuit behaves as a pure resistance $R/2$ at all frequencies.

CHAPTER III

CIRCUIT ELEMENTS

1. The Use of Equivalent Circuits to Describe the Electrical Properties of Circuit Elements.—The physical elements that comprise the electric circuit possess a mixture of capacitive, inductive, and resistive characteristics. Although the coil of wire that constitutes an inductor has a geometry and construction designed to make inductance its primary attribute, there are always resistance and capacitance associated with it. Similarly a stack of plates forming a capacitor has inductance and resistance present as residual quantities.

The influence of the element upon the electric circuit depends upon the current-voltage relations it maintains at its terminals. Over a range of frequencies these relations change in a complicated fashion depending upon the geometry of the element and the distribution in space of its electric and magnetic fields. At any one frequency the ratio of the voltage at the terminals to the current at the terminals may be described by stating the impedance or admittance of the element, or by attributing to the element the properties of an equivalent series or parallel circuit (of L and R or C and R) having the same value of Z or Y as has the element at the frequency in question.

If the frequency is varied, it is necessary in general to vary the magnitudes of L , R , or C , in this simple equivalent circuit if the calculated Z is to agree with the actual Z of the circuit element. By resorting to a more complicated arrangement of ideal elements of constant L , R , C , it is possible to devise an equivalent circuit whose impedance variation over a frequency interval is similar to that of the actual element. This agreement can never be perfect over the entire range of frequencies; and, the wider the range and the greater the accuracy required, the more complex the equivalent circuit must be.

In general, the simplest possible circuit is used that yields sufficient agreement with the behavior of the circuit element over the range of frequencies of interest. For an air-core inductor near zero frequency the equivalent circuit may be reduced to a simple

series combination of resistance and inductance of constant magnitudes. As the frequency is increased, the effect of the distributed capacitance causes the equivalent series resistance and inductance to change in a manner that can be accounted for by including a capacitor in the equivalent circuit. At still higher frequencies, however, the current distribution in the circuit element changes greatly from its low-frequency pattern, and the "constants" of the equivalent circuit again must be altered.

To be able to predict the behavior of a circuit element by the use of alternating-current circuit theory, the establishment of an equivalent circuit is necessary. As the nature of the equivalent circuit is related to the geometry and construction of the circuit element, the equivalent circuit may serve as a guide in designing the element for special purposes.

2. Electrical Representation of an Isolated Inductor or Resistor.

An inductor or a resistor consists of a conductor of such material

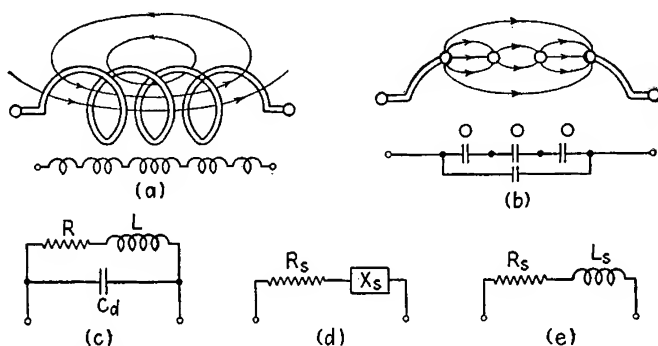


FIG. 2.1.—Distributed magnetic and electric characteristics of a coil and their representation as lumped elements.

and of such geometrical form as to cause inductance or resistance to be its predominant electrical characteristic. One simple arrangement used for both elements is that of a single-layer solenoid, Fig. 2.1a. In addition to the magnetic field that accompanies an alternating current in the element, there will exist also an electric field due to the voltage across the element. An axial section of the coil is indicated in Fig. 2.1b. Owing to the momentary excess of positive charge at the + end of the coil over that at the - end, lines of force are established in the manner indicated in the figure. One half period later the polarity of the charge and the direction of the lines of force are reversed. The instantaneous surface-

charge density varies along the coil from a maximum positive at one end through zero at the center to a maximum negative at the other end; and as this distribution changes periodically, a charging current similar to the charging current of a capacitor is supplied at the terminals of the coil. The presence of this current is accounted for in Fig. 2.1c by a lumped capacitance C_d between the terminals, shunting the lumped resistance and inductance of the coil. The magnitude of C_d depends upon the diameter of the coil and other factors, including the distance between turns and the length of the coil. When the ratio of the diameter of the solenoid to its length does not differ greatly from unity, C_d in micromicrofarads is of the order of magnitude of one-quarter of the diameter of the coil in centimeters (the effects of any dielectric other than air being neglected).

The impedance variation of the coil, with its mixed inductive, capacitive, and resistive characteristics, can be closely matched over a range of frequencies by the circuit of Fig. 2.1c made up of pure lumped constants L , R , C . The expression for the impedance of this parallel combination is calculated in Chap. I, Sec. 13. The variation of its impedance with frequency predicted there by (13.6) is more readily followed if Z is expressed in terms of the quality factor of the series loop of Fig. 2.1c ($Q_r = \frac{1}{R} \sqrt{\frac{L}{C_d}}$) and the ratio of the exciting frequency to the series-resonant frequency of the loop ($w = \omega/\omega_{sr} = \omega \sqrt{LC_d}$).

$$Z = R \frac{1}{(1 - w^2)^2 + \frac{w^2}{Q_r^2}} + j\omega L \frac{(1 - w^2) - \frac{1}{Q_r^2}}{(1 - w^2)^2 + \frac{w^2}{Q_r^2}} \quad (2.1)$$

$$= R_s + jX_s = R_s + j\omega L_s \quad (2.2)$$

These equations justify the alternative equivalent series circuits of Fig. 2.1d and Fig. 2.1e, where R_s , X_s , L_s are functions of frequency. The predicted variations of Z are quite different in character depending upon whether the values of L , C , R are chosen of proper relative magnitudes to describe an inductor or a resistor.

3. Frequency Dependence of Inductor Characteristics.—It is important to distinguish between the variations of Z predicted by the circuit of Fig. 2.1c and the impedance variations of an actual inductance coil. For such a coil R and C_d are small residual quantities in the equivalent circuit so that Q_r in (2.1) is always

large. The terms involving $1/Q_r^2$ in (2.1) may then be neglected for values of w ranging from zero up to nearly unity.

$$Z = R_s + jX_s \doteq R \frac{R}{(1 - w^2)^2} + j\omega L \frac{1}{(1 - w^2)} \quad (3.1)$$

The equivalent series resistance and reactance of the circuit are given by

$$R_s \doteq \frac{R}{(1 - w^2)^2} = \frac{R}{(1 - \omega^2 LC_d)^2} \quad (3.2)$$

and

$$X_s = \omega L_s \doteq \frac{\omega L}{1 - w^2} = \frac{\omega L}{1 - \omega^2 LC_d} \quad (3.3)$$

If the frequency is reduced toward zero, it is apparent from (3.2) and (3.3) that R_s and L_s approach the values R and L . If the equivalent circuit is to match the characteristics of the actual coil in this range of frequencies, R and L must be taken as equal to the measured d-c or low-frequency value of the coil resistance and self-inductance, designated as R_{dc} and L_{dc} . The calculated equivalent series resistance and inductance may be compared with these constant values. Similarly, the quality factor Q_s of the equivalent series circuit may be expressed in terms of Q_r , the quality factor of the series loop.

$$\frac{R_s}{R_{dc}} \doteq \frac{1}{(1 - w^2)^2} \quad \frac{L_s}{L_{dc}} \doteq \frac{1}{1 - w^2} \quad Q_s = \frac{X_s}{R_s} \doteq Q_r(w - w^3) \quad (3.4)$$

These relations for a low-loss coil are independent of Q_r . They are valid provided that the frequency does not rise too near to ω_{sr} , the resonant frequency of the series loop. In the neighborhood of ω_{sr} the ratio w approaches unity, R_s approaches a maximum value $Q_r^2 R$, and L_s drops abruptly to zero. The ratios of (3.4) are illustrated in Fig. 3.1a. It is seen that R_s increases more rapidly than L_s . The quality factor Q_s of the coil predicted by this circuit with constant elements L_{dc} , R_{dc} , C_d has a variation as shown in Fig. 3.1b. Increasing at first linearly with frequency, it approaches a broad maximum of $0.385Q_r$ at a frequency of $0.578/\sqrt{3}$ times the resonant loop frequency. Above this frequency Q_s drops rapidly toward zero.

The actual (measured) values of the series resistance R_a , inductance L_a , and quality factor $Q_a = \omega L_a/R_a$ for an inductor may depart markedly from these ideal values, which assume the L , R , C_d of the

coil to be constant. The dashed curves of Fig. 3.1 indicate the actual variations for a single-layer air-core coil having its maximum Q_a in the broadcast range. The most striking deviation from the ideal curves is the rapid increase of R_a , due mainly to an increase in resistance associated with the skin effect. The value of L_a is reduced as compared with the ideal case, and these two changes combine to give a greatly reduced value of Q_a with a broadened maximum.

The representative circuit of Fig. 2.1c has a maximum impedance almost exactly at its resonant frequency, and at this frequency the reactance X_s changes from inductive to capacitive, as shown in Chap. II, Fig. 14.1. In the frequency range near resonance the

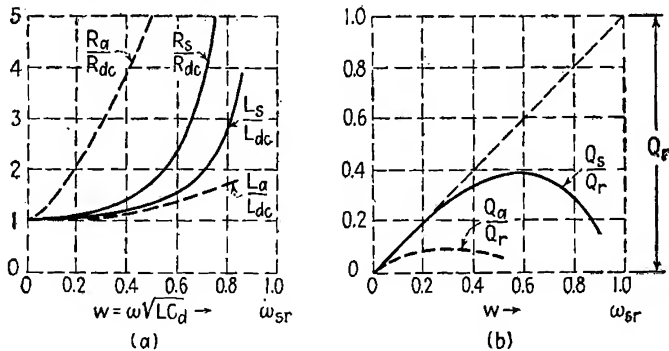


FIG. 3.1.—Computed variations of the series L , R , and Q for the equivalent circuit compared with the measured values for an actual coil.

current becomes so nonuniform along the coil, owing to the distributed capacitance, that the equivalent circuit is useful only as a general guide to the electrical characteristics of the element. The resonant angular frequency of the loop ($1/\sqrt{L_{dc}C_d}$) still allows an estimate to be made of the frequency at which the series inductance of the coil becomes zero. This calculated value is generally too low. It ranges between 60 to 80 per cent of the actual self-resonant frequency of the isolated coil.

The calculated behavior of the equivalent circuit indicates the importance of controlling C_d in an actual coil. The self-resonant frequency of the coil sets an upper limit to the frequency at which it is useful in providing an inductive reactance. In the design of coils to be used at high radio frequencies it is important to reduce C_d . Such coils commonly consist of a comparatively small number of turns of bare wire in the form of a single-layer helix. The wires

may be large enough to be self-supporting and are generally spaced from one to two or more wire diameters apart. The absence of a coil form with a large dielectric constant and the spacing of the wires both reduce C_d .

In coils of larger inductance it is necessary to use several layers of insulated wire. The distributed capacitance is minimized by the choice of an insulation between wires that has a low relative dielectric constant, by spacing the wires where possible, and by keeping the potential difference between adjacent wires low. If the second layer of wire is wound back over the first, the average potential difference between wires in the two layers is equal to the drop across one layer. At the expense of greater accuracy in winding, the wires can be made self-supporting and a minimum potential difference obtained between adjacent wires by the use of bank winding as in Fig. 3.2*a*. The winding is completed in one motion along the

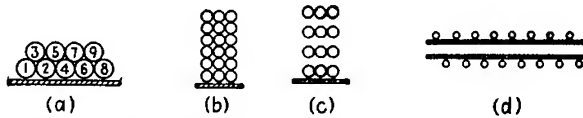


FIG. 3.2.—Types of winding used to reduce distributed capacitance of inductors.

length of the coil with successive turns laid on in the order indicated. Other self-supporting windings are made by feeding the wire back and forth along the tube at a large pitch angle. The resulting coil is porous and the average separation between wires increased. The potential between adjacent wires can be reduced by keeping the coil short in comparison with the thickness of winding, Fig. 3.2*b*. By spacing the windings as in Fig. 3.2*c* the capacitance is further reduced. Radio-frequency choke coils are often made of several short coils similar to those shown in Fig. 3.2*b* arranged in series and spaced along an insulating tube. If they are operated at a fixed frequency, the increase in L_s somewhat below their self-resonant frequency may be utilized to make them more effective. If the frequencies applied cover the range of self-resonance and if the coils do not have the same self-resonant frequency, not only is it possible to have the unit appear capacitive, but owing to series resonance between the equivalent circuits of the separate coils the impedance of the combination may drop to a low value at some frequencies. At very high frequencies, a long coil of small diameter with widely spaced turns is sometimes used as a radio-frequency choke, Fig. 3.2*d*.

By the use of magnetic core materials, coils of large inductance can be made with a greatly reduced number of turns and a corresponding decrease in d-c resistance and physical size. The eddy currents induced in a solid core reduce the self-inductance and increase the energy loss of the coil. At low frequencies this is avoided by making the core of thin laminations insulated from one another. As the frequency is increased, the core material must be still more finely divided. By using powdered magnetic particles immersed in an insulating medium it is possible to extend the use of magnetic cores into the intermediate- and radio-frequency range. The losses of such a coil, which increase its equivalent series resistance, depend upon the size, resistivity, and hysteresis of the magnetic dust particles as well as the dielectric characteristics of the insulating medium. The losses and the value of the self-inductance are also dependent upon the frequency and magnitude of the current in the coil.

4. Frequency Dependence of Resistor Characteristics.—In a resistor the inductance and capacitance are residual quantities, and L and C_d in the equivalent circuit of Fig. 2.1c are small enough to cause a low quality factor for the circuit. The value of Q_r may be either larger or smaller than unity, depending upon the magnitude of the resistance and the method of winding. Resistors of low value (say up to 100 ohms) generally have L/C ratios yielding values of Q_r greater than 1. Reference to Fig. 15.2 of Chap. II shows that for this case both R_s and X_s increase with frequency in the low-frequency range. For larger values of resistance, Q_r decreases and may become less than unity, in which case X_s is always capacitive and R_s may increase or decrease with frequency.

The two essential characteristics of a pure resistance are its zero phase angle and the constant magnitude of its impedance with varying frequency. With a circuit similar to Fig. 2.1d these characteristics can be maintained over a range of frequencies by reducing L and C_d to very low values. If, in addition, the ratio L/C_d is adjusted properly, it is possible to obtain a close approximation to one or the other of these characteristics of a pure resistance over a greatly increased range of frequencies. If the ratio L/C_d is adjusted to make $Q_r = 1$, the variation in phase angle has its minimum value, Fig. 16.1a of Chap. II. For a Q_r of 0.707, R_s is nearly constant, Fig. 16.1b of Chap. II, as shown by the following approximation from (2.1), where w^4 is neglected.

$$R_s \doteq \frac{R}{1 - w^2 \left(2 - \frac{1}{Q_r^2} \right)} = \frac{R}{1 - \omega^2 C_d (2L - R^2 C_d)} \quad (4.1)$$

With L/C_d such that $Q_r = 0.707$, $|Z|$ increases only a few per cent above its zero-frequency value as the frequency is increased and again returns to its zero-frequency value at a frequency that is 70.7 per cent of the resonant frequency of the loop.

Again the equivalent circuit serves as a guide in the construction of a resistor of desired characteristics. It is important not only to reduce L and C_d for high-frequency operation but also to control their ratio. This is accomplished by the use of the proper form of winding. If a straight section of wire is bent back upon itself in the form of a hairpin, the magnetic effects of the opposing currents tend to cancel each other and the self-inductance is reduced. Such a winding is described as bifilar and is one of the simplest noninductive windings. It is generally used for low-resistance units, where the L/C_d ratio resulting from this winding yields a value of Q_r near unity. The more closely the wires approach each other, the smaller the area of the hairpin loop and the smaller the self-inductance. However, the distributed capacitance between the wires is correspondingly increased, and by choosing a suitable separation the L/C_d ratio can be controlled.

The wires can be placed close together and the value of C_d prevented from becoming large by separating the winding into sections. Consider the same length of resistance wire arranged as shown in the three sketches of Fig. 4.1. The separation of the wires is assumed small in comparison with their length. The electric lines of force shown indicate the relative nature of the charge distribution when equal potential differences are maintained across the terminals of all three circuits. In Fig. 4.1a the wire is cut at the center, and the potential difference is constant along the wire, as is also the charge distribution. The lumped capacitance representing that of the wire is the ratio of the charge to the applied terminal potential. If the far end of the wire is shorted as in Fig. 4.1b, the potential difference decreases along the wire, approaching zero at the far end, and the average potential difference between the wires is $\frac{1}{2}$ its previous value. The charge stored and the effective value of C_d for the first case is therefore twice that for the second case, which represents a single-section bifilar resistor. If the same

wire is divided as in Fig. 4.1c into two bifilar sections, the average potential difference between adjacent wires and the length of each section is $\frac{1}{2}$ that of Fig. 4.1b. If the wires are separated enough so that practically no field exists between sections, the charge supplied per section, which is the charge supplied at the terminals, is $\frac{1}{4}$ that of the original bifilar winding. Therefore the value of C_d is reduced to approximately $\frac{1}{4}$ of the single-section value by dividing the bifilar element into two sections; to $\frac{1}{8}$ for three sections; etc. The reduction occurs at the expense of increasing the self-inductance slightly because the short length of conductor joining two sections is uncompensated.

The sectionalized bifilar winding illustrates the principles employed in general to reduce the effect of distributed inductance and capacitance and to control their relative magnitudes. The methods used have as their purpose to cancel the magnetic effect

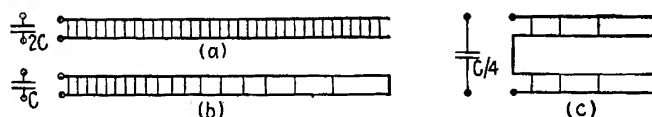


FIG. 4.1.—Reduction of C_d by arranging a bifilar winding in sections.

of any current by an equal, oppositely directed current near it and to reduce the value of C_d by keeping the potential difference between adjacent conductors as small as possible. These precautions reduce the magnetic and electric fields and the stored energy associated with them.

Many types of winding are used for the control of residual inductance and capacitance. Their nature depends upon the magnitude of the resistance and the frequency range for which they are designed. A simple type of winding consists of a series of adjacent turns wound like a flattened coil around a thin sheet or card of insulating material. Its inductance is small due to cancellation of the magnetic effects of the currents on one side of the sheet by the oppositely directed currents on the other side. Owing to the large number of turns the potential difference between adjacent turns is small, and therefore C_d is not large. The residual inductance can be reduced further by using two windings in parallel, wound in opposite directions around the card. This type of winding is called the Ayrton-Perry winding. Resistance wire and insulating fibers may be woven into a tape in several patterns designed to reduce inductance and distributed capacitance. In

wire-wound resistors the wire may be wound on a ceramic spool with adjacent grooves connected by a longitudinal slot. The separate coils in the spaced grooves provide a sectionalization that reduces C_d , and the winding is reversed in adjacent grooves to decrease the inductance. In the "fishline" winding, used for high resistances, fine resistance wire is wound in a spiral of very small diameter around an insulating cord. This cord may then be wound on a coil form or in any of the patterns described.

An alternative to the use of long lengths of metal wires is the use of much shorter lengths of materials of higher resistivity. These materials may consist of a finely divided conductor, such as colloidal graphite, dispersed in some inert insulating medium. As compared with pure metal resistors, the resistance of these elements is more dependent upon time, temperature, and the magnitude of the applied voltage. The variation of resistance with time may be of short period and associated with random changes in the internal "contacts" of the finely divided conductor. If the resistor is used in the first stage of a high-gain amplifier, this erratic change in resistance produces "noise." The noise introduced varies from resistor to resistor even of the same kind and may be reduced by careful selection. Long-period changes are due to aging of the resistor or to temperature changes. The resistance in general decreases with increased temperature and applied voltage. These disadvantages of composition resistors are outweighed for many purposes by the low values of L and C_d associated with them, which make them approximate a pure resistance over a wider range of frequencies. Their reduced size also results in a smaller capacitance to ground. Owing to their high resistivity and simple form the skin effect (see Sec. 5) is small as compared with wire-wound resistors.

Such composition resistors may be produced in several forms. In some cases the resistor consists of a short length of the semiconducting composition with leads attached. In others the conducting composition forms a small core embedded in a thick protective insulating layer, which also serves to immobilize the lead wires and their connection to the core. In some units the current is carried by a thin coating of semiconductor on an insulating cylinder of small diameter, which is also protected by an outer layer of insulation.

The wattage rating of a resistor is a measure of its ability to dissipate heat under specified conditions without a resulting temperature increase sufficient to damage it or affect its resistance characteristics. The rating depends upon the dissipating surface

and its radiating properties as well as the thermal properties of the materials used and their ability to withstand high temperatures. It follows that the physical dimensions of a resistor of given construction are determined by its wattage rating and do not depend upon its resistance. The wattage rating selected for a resistor mounted in a confined space where free motion of air is impossible or near other heat-producing elements should be two or three times the normal value. In general, an underloaded element provides more stable operation and longer life.

5. Skin Effect.—As the frequency is increased, the current flow across a section of a conductor changes from the uniform distribution

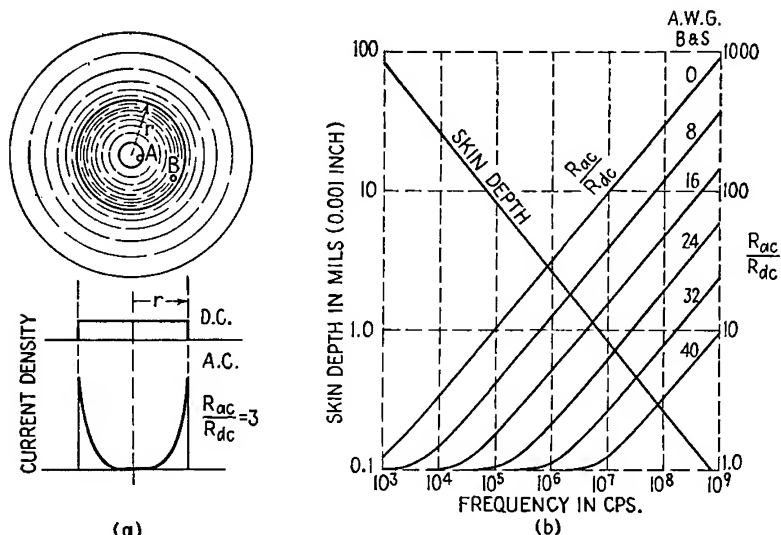


FIG. 5.1.—(a) Nonuniform current density due to skin effect; (b) skin depth and resistance ratios due to skin effect.

of direct current to one in which an increasing portion of the current flows near the surface of the conductor. At sufficiently high frequencies the current may be confined to a thin layer at the surface. This is known as the *skin effect* and causes the high-frequency resistance of the conductor to be larger than its d-c resistance, by a factor that is larger the greater the current concentration.

The reason for the concentration and the factors controlling it are illustrated by consideration of an isolated straight conductor of circular section carrying a direct current. The magnetic field of the conductor is shown qualitatively in Fig. 5.1a. If the current is

decreased to zero, the magnetic field collapses and induces a voltage in the conductor that tends to oppose the changing current. The voltage induced is not uniform over the conductor section; for while most of the lines of force cut the point *A* near the center, those inside of the point *B* near the circumference do not cut it upon collapse. Therefore a sinusoidally varying current, uniform over the cross section, would cause a larger self-induced voltage near the center than near the periphery of the conductor. The higher the frequency and the greater the diameter and permeability of the conductor, the larger the difference in voltage. The greater the conductivity of the conductor, the larger the redistribution of current due to this induced voltage differential. The corresponding redistribution of current and change in resistance for such a conductor are determined by the factor $d\sqrt{\mu\sigma f}$, where d , μ , and σ are the diameter, permeability, and conductivity of the conductor and f is the frequency of the current.

At low frequencies where the change in resistance is small, the redistribution of the current is slight, and the *per cent change* in resistance is proportional to the fourth power of the above factor, or the change is directly proportional to the square of the area, permeability, and conductivity of the conductor and to the square of the frequency. At higher frequencies the current is concentrated near the surface with a negligible value near the center. As the central part of the conductor is not used, the resistance of the conductor is increased. This is indicated in Fig. 5.1*a*, where the uniform current density over a conductor section that exists for sufficiently low frequencies is compared qualitatively with the distribution at a higher frequency producing an approximate threefold increase in resistance. When the current is concentrated in a layer whose thickness is small in comparison with the diameter, it is possible to establish a "skin depth." This is equal to the thickness of a hollow cylinder, of the same material and having the same outside diameter as the wire, whose d-c resistance is equal to the high-frequency resistance of the solid conductor. For values of skin depth such that the ratio of a-c resistance to d-c resistance increases to 3 or more, the ratio R_{ac}/R_{dc} becomes proportional to $d\sqrt{\mu\sigma f}$, or varies as the square root of the frequency. As the d-c resistance of the wire is inversely proportional to d^2 and R_{ac}/R_{dc} is proportional to d , the high-frequency resistance is inversely proportional to d (or the perimeter of the wire under these circumstances). This may be extended to conductors of any shape

provided that the skin depth is quite small in comparison with the dimensions of the cross section and the maximum radius of curvature of its surface. Under these conditions the resistance of the conductor is inversely proportional to its perimeter, and its high-frequency resistance may be calculated in terms of the skin depth, perimeter, resistivity, and length. Values of skin depth for copper under these conditions are shown as a function of frequency in Fig. 5.1b. They may be extended to other metals by noting that the skin depth is proportional to the square root of the conductivity. The variation of R_{ac}/R_{dc} is also shown for straight, isolated, round copper wires of American wire gauge Nos. 0 to 40.

If the section of the conductor is other than circular, the current flows in such a manner as to link a minimum number of lines of force. In a rectangular bar or strip this involves a current concentration at the corners or edges. The per cent change in resistance from the d-c value is decreased for a circular conductor by removing the central portion, yielding a tubular conductor with a minimum current redistribution with frequency. By reducing the diameter of the wire the per cent change in resistance also is decreased. If a conductor is made up of fine insulated wires, so transposed along its length that each wire appears at the surface to the same extent as any other, equal currents will flow through the wires and the change in resistance is minimized. Such a conductor is called Litzendraht or *litz* wire. There always exists an upper frequency limit at which its advantages are lost owing to the current flowing by capacitive action between the wires at the surface. Depending upon its design it is useful up to intermediate or lower radio frequencies.

When two conductors carrying alternating currents are parallel, there will be a still further redistribution of current owing to their interaction. In general, this results in a further increase in the resistance. This is known as the *proximity effect*, and the nature of the new current distribution depends upon the relative direction of the currents. If the currents are in the same direction, the current shifts so as to increase the distance between the major portions of the current. For two circular conductors this further increases the R_{ac}/R_{dc} ratio, but by a moderate amount. If the currents are in opposite directions, the shift is toward the adjacent portions of the conductors; and if the separation is of the order of a fraction of a wire diameter, the increase in resistance may be manyfold. These effects are minimized in r-f coils by spacing adjacent wires and

using single-layer solenoids where possible. For multiple-layer coils the proximity effect is an important factor affecting their h-f resistance.

The redistribution of current due to the skin and proximity effects reduces the self-inductance of a wire or coil slightly and causes a still smaller change in C_d . In many cases these changes can be ignored in comparison with the major change in resistance.

6. Electrical Representation of a Capacitor.—Inductance and resistance appear as residual characteristics in a capacitor. The inductance, which is a series element, depends upon the arrangement of the leads and the connections to the capacitor plates. There likewise is a series resistance due to the leads and also due to ohmic losses in the capacitor plates as the charging current flows through them. The resistance in the leads and in the capacitor

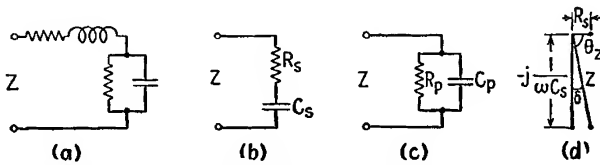


FIG. 6.1.—Representative circuits for a capacitor.

plates may be increased by the skin effect at high frequencies. There also may be a leakage current through the capacitor, properly represented by a resistor in parallel with C . A possible representative circuit for the capacitor is that of Fig. 6.1a. The series inductance and resistance are of increasing importance at very high frequencies, where the capacitive reactance becomes small, and every precaution must be taken to reduce them in a capacitor designed for high-frequency use.

At low frequencies the self-inductance is commonly neglected and the dissipation of the capacitor is represented by a single resistance either in series or in parallel, Fig. 6.1b or c. The relations between the constants in such alternative series or parallel representations of a dissipative reactor have been established in (14.7), Chap. I. The losses in a well-made capacitor are so small that the approximations of (14.8), Chap. I, apply. The relations between R_s , C_s and R_p , C_p become

$$C_p \doteq C_s \quad \text{and} \quad R_p \doteq Q^2 R_s \quad \text{where} \quad (6.1)$$

$$Q = \frac{X_s}{R_s} = \frac{1}{\omega C_s R_s} = \omega C_p R_p = \frac{1}{D}$$

Q is the quality factor of the capacitor and its reciprocal D serves as a measure of its losses. If Z , Fig. 6.1*d*, represents the impedance of the capacitor at some frequency, the loss angle δ is also a measure of the losses of the capacitor. As the losses decrease, δ becomes small and $|Z|$ approximates $1/\omega C_s$. This allows the following relations to be established among the dissipation factor, the loss angle, and the power factor of the capacitor:

$$D = \frac{R_s}{1/\omega C_s} = \tan \delta \doteq \delta_{\text{radians}} \doteq \cos \theta_z = \text{the power factor} \quad (6.2)$$

7. Capacitor Characteristics.—The important characteristics of a capacitor are its capacitance, maximum voltage rating, and losses. All these characteristics depend upon the geometry of the capacitor and the nature of the dielectric. All may vary somewhat because of changes in temperature or frequency.

The capacitance is proportional to the area and inversely proportional to the separation of the plates. It also varies directly with the relative dielectric constant ϵ_r of the dielectric. This constant is determined experimentally as the ratio of the capacitance of a capacitor when the electric field is established in the dielectric as compared with its capacitance with air or vacuum surrounding the plates.

The maximum voltage that may be applied safely to a capacitor depends upon the dielectric, the separation of the plates, and their shape. Irregularities in their surface or sharp corners tend to produce concentration of the lines of force and increased potential gradients tending to break down the dielectric. The dielectric affects the breakdown voltage of a capacitor both because of the limited constant potential gradient it can withstand and because of its loss characteristics.

For most well-designed capacitors the losses due to conductor resistance are negligible in comparison with those within the dielectric. If an alternating voltage were applied to an ideal capacitor, the energy supplied as the capacitor is charged would be completely returned to the circuit upon discharge one-half cycle later. Practically, there is a difference between the two energies representing a loss or dissipation that must be explained in terms of resistance.

It is mainly losses of this sort that are described by the resistances in the representative circuits of Fig. 6.1*b* or *c*. The parallel combination of a constant C_p and R_p simulates the actual variations

of Z for most capacitors over a wider range of frequencies than the equivalent series circuit. However, the series representation is commonly used, in which case R_s decreases with increasing frequency.

The loss characteristics of a dielectric material may be described in terms of the loss angle δ of a capacitor whose plates are immersed in the material, it being assumed that the dielectric is responsible for all the losses. The angle δ may be expressed in degrees or radians or in terms of its tangent.

8. Types of Capacitors.—As the characteristics of capacitors are so dependent upon the dielectric used, it is convenient to discuss gaseous, liquid, and solid capacitors separately.

Air, or Gaseous Capacitors.—Capacitors with air as a dielectric are characterized by low losses and small values of capacitance seldom over a few hundred micromicrofarads. Although comparatively bulky, they can easily be made variable, and with proper design their capacitance at any setting is accurately reproducible and nearly independent of frequency.

The voltage breakdown is generally due to the establishment of a corona discharge. This occurs when the voltage gradient near some sharply curved surface becomes sufficient to produce ionization of the gas and a glow discharge. This frequently occurs at the edges of the plates, and capacitors designed for high-voltage applications generally have thick, widely separated plates with rounded edges to reduce the voltage gradient.

The breakdown voltage is markedly dependent upon the pressure of the gas surrounding the plates. At the very low pressures obtainable with modern vacuum technique (say 10^{-9} atmosphere) the corona effect is eliminated, and the breakdown voltage is greatly increased. Such vacuum-type glass-enclosed capacitors give a compact unit of high voltage rating. With a poorer vacuum (higher pressure) the breakdown voltage decreases and reaches a minimum at pressures of the order of magnitude of $\frac{1}{100}$ atm. With further increase in pressure the breakdown voltage rises, and at pressures of several atmospheres it is increased several fold as compared with its value at atmospheric pressure. Such high pressures are frequently employed to increase the voltage rating of equipment. A corresponding reduction of voltage rating occurs when electrical equipment is flown to high altitudes. At the lower barometric pressure corresponding to an altitude of 5 miles the breakdown voltage of an air capacitor is reduced to approximately one-half its value at the earth's surface.

At moderate frequencies most of the loss associated with air capacitors occurs in the insulation used to maintain the separation of their plates. This loss is minimized by placing the insulation where the electric field is weak and choosing its shape so that it is long and narrow in the direction of the electric field. As contrasted to materials used for dielectrics between the plates of liquid- or solid-filled capacitors this insulating material should have a small value of ϵ_r to minimize its effect of increasing the capacitance and of course should have a small value of $\tan \delta$. The product of ϵ_r and δ is used commonly as a figure of merit for such materials. The insulation is generally placed where the strength of the field and therefore the losses are independent of the position of the movable plates which determine the capacitance.

Liquid-filled Capacitors.—The introduction of a liquid dielectric between the plates of an air capacitor may increase its capacitance several fold and also may increase its breakdown voltage. The losses are also increased, and in general both the loss angle and the dielectric constant depend upon frequency.

In a liquid dielectric the alternating electric field causes motions of charged ions, rotations of polar molecules, and molecular distortions calling for a supply of energy during one part of the cycle that is not completely returned to the circuit during the remainder of the cycle. Liquids with unusually large dielectric constants have in general large values of δ .

Solid-dielectric Capacitors.—Solid dielectrics frequently used in capacitors are mica and wax- or oil-impregnated paper. Mica capacitors have comparatively large voltage ratings but are generally of low capacitance, up to about $0.02 \mu f$, as compared with values of 6 or $8 \mu f$ available in paper units. They also have much lower losses than paper capacitors, particularly at radio frequencies. Mica capacitors consist generally of a pile of alternate sheets of metal foil and mica held together by some pressure device. Paper capacitors consist of long strips of foil and specially treated paper rolled together in a compact unit. Both types are enclosed in a protective cover, which should be airtight.

The capacitance of such units may vary slowly as the pressure maintaining the foil separation changes with time. Temperature changes also change the value of the capacitance by changing the capacitor dimensions and the dielectric characteristics. Capacitors of quite stable value can be made by depositing a metal film directly on the surface of the mica sheets. By using certain ceramic

materials whose dielectric constant increases or decreases with temperature it is possible to obtain a capacitor whose temperature coefficient cancels the effect of temperature variation in other parts of an oscillatory circuit. By the use of such capacitors, oscillators may be constructed whose frequency is almost independent of temperature.

The voltage rating of a solid-dielectric capacitor decreases with frequency. At low frequencies it is determined by the maximum potential gradient the dielectric will withstand without rupture. As the frequency is increased, the increased dielectric losses and the resulting temperature rise become the limiting factors, and the voltage rating may be reduced by a factor of 100 or more. The corresponding rated current increases from low values at low frequencies to an upper limit at high frequencies.

9. Electrolytic Capacitors.—The breakdown voltage of a capacitor is determined by the weakest point in its dielectric. The

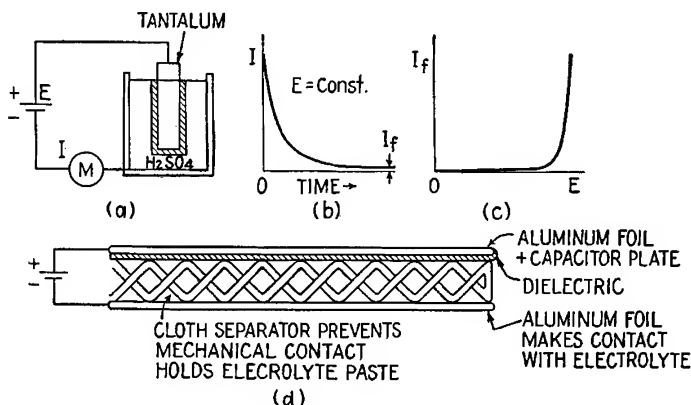


FIG. 9.1.—Simplified pictures of electrolytic capacitors.

dielectrics mentioned thus far cannot be made too thin without increasing the possibility of holes or large percentage variations in thickness. By chemical means it is possible to establish uniform insulating layers whose thickness may be a very small fraction of a thousandth of an inch.

If a tantalum sheet, immersed in sulphuric acid of about the concentration used in storage cells, is made the positive plate in an electrolytic cell, Fig. 9.1a, the large initial current decreases with time to a small final value I_f , Fig. 9.1b. This is due to the formation of an insulating layer over the surface of the tantalum.

Any break or thin spot in this layer causes a current concentration, which builds up the insulation to a value uniform over the surface. The insulation is therefore self-healing and self-adjusting in thickness. Thus there is formed a capacitor having the tantalum as one plate, the insulating layer as a dielectric, and the electrolyte as the other plate. The thickness of the layer increases with the voltage applied, as does the final leakage current. This is illustrated in Fig. 9.1c, which shows that there is an upper limit of voltage above which the leakage current rises rapidly. If the voltage is removed, the layer gradually dissolves and in this type of cell would have to be re-formed upon the next application of voltage. Other metals and electrolytes can be used; this so-called "wet" type of electrolytic capacitor with several modifications is mainly used in permanent installations.

The most commonly used electrolytic capacitors are of the semi-dry or dry type, which may consist of two sheets of aluminum foil between which is held an electrolyte that has been thickened to a paste or even a solid at ordinary temperatures. The strips of foil are separated by a porous cloth impregnated with the electrolyte. If a battery is connected as shown in Fig. 9.1d, the positive aluminum sheet develops an insulating layer as before, and a capacitor is formed with the electrolyte serving as the other plate. The second aluminum foil serves only to give electrical contact with the electrolyte. In practice, the positive foil has a coating, or film, formed on it in a previous operation by a voltage which determines the voltage rating. This film may be made quite permanent, and the action of the electrolyte in the finished capacitor is mainly to repair and maintain the originally formed film. The two foils and the separator are rolled together and sealed in a container.

The capacitance of the unit depends upon the surface area of the positive foil (which may be increased by embossing or etching it) and the forming voltage, which determines its voltage rating. The larger the forming voltage, the thicker the insulating film and therefore the smaller the capacitance. For a given area of foil the product of voltage rating and capacitance is approximately a constant above a forming voltage of about 100 volts. The capacitance is increased slightly and the leakage current strongly by an increase in temperature. The capacitance decreases with frequency and may fall off markedly in the upper audio range.

The losses in an electrolytic capacitor are large in comparison

with those of the other types studied, corresponding to values of $D (= \tan \delta)$ of the order of 0.1. They are due to dielectric losses and leakage, similar to those of other capacitors but larger; in addition, there are losses due to the resistance of the electrolyte. Though it is difficult to separate their effects this last is probably the principal loss.

The polarity of electrolytic capacitors should be carefully observed when they are installed, as the film has rectifierlike characteristics and therefore a lower resistance in the reverse direction. With time the film is partly dissolved by a reversed voltage, and the corresponding film formed on the other aluminum foil does not have the desirable characteristics of the original one. The ripple component should not be too large because of the possibility of reversing the polarity momentarily and also because its heating effect may cause a rapid deterioration.

Several capacitor units may be assembled in the same container, frequently sharing a common negative aluminum strip. The units are seldom made with voltage ratings much above 550, but occasionally capacitors of larger voltage ratings are made with two such units in series. If two such formed units were connected in series with the polarities opposing, they would serve as a capacitor for use in a-c circuits, with the separate capacitors serving to block the current every other half cycle. The same action can be obtained by preforming a hard insulating layer on each aluminum foil in a single unit. Such capacitors are used in motor-starting systems or other applications where an a-c capacitor of comparatively large losses can be used.

10. Effects of Moisture.—The presence of moisture has an important effect upon the characteristics of most electrical elements. Even under conditions of moderate humidity an invisible thin film of water forms on the surface of most substances and is absorbed into their interior if they are at all porous. Although pure water has a high volume resistance, the small amount of dissolved impurities present makes it possible to cause changes of a millionfold in the surface and volume leakage of insulators. Owing to its large dielectric constant and large value of $\tan \delta$, it also introduces important dielectric losses if it is exposed to an alternating electric field.

In the manufacture of capacitors, this is recognized by the efforts made to drive out water vapor and to impregnate and surround the capacitor by an insulating compound generally applied

at an elevated temperature. Impregnation under partial vacuum reduces or eliminates entrapped air bubbles together with their water-vapor content and does away with their action in producing high local electric stresses. As a further precaution the unit may be enclosed in a sealed container to prevent slow penetration of water vapor.

Similar precautions should be taken in the construction of inductors, particularly if they are to be exposed to high-humidity conditions. Under conditions of low humidity the dielectric losses due to the insulation between turns and in the supporting form may be made negligible up to very high frequencies by suitable choice of materials. With the presence of water vapor on the surface or porous interior of the insulation there is an increase in C_d due to the large dielectric constant of water and also a marked increase in the losses in the coil. This possibility can be reduced by impregnating the coil with a thick layer of water-repellent wax.

Erratic behavior of equipment may sometimes be traced to the failure of insulators due to decrease of surface or volume resistivity caused by water. This may come slowly after a few days of high humidity or abruptly because of a change in temperature that causes condensation of water droplets, which unite to form a conducting path over the surface. The droplets can be driven off or prevented by maintaining an elevated temperature in the equipment or may be controlled by covering the insulation with a surface that is not wetted by water so that the droplets remain separate. Prevention of the slow penetration of water into fine-grained porous insulators such as mica and some ceramic materials requires thick and unbroken coatings. Once water is absorbed, it is driven from the interior only with the greatest difficulty.

11. Shielding.—For simplicity it has been assumed in earlier sections that the circuit element studied was isolated. The action of such an element in an electric circuit, however, depends upon the nature of the surroundings to the extent that its electric and magnetic fields affect them. This interaction can be reduced by shielding, which at the same time decreases undesired induced voltages and renders the impedance of the shielded element more definite.

The effect of adjacent objects upon the electrical characteristics of an element is of increasing importance at high frequencies and depends in a complicated manner upon the geometry and the potentials involved. In many cases, however, the effect of the

surroundings can be described electrically by altering slightly the representative electric circuit of the isolated element.

Let a uniform circuit element, which may be a resistor or inductor, lie along a line between the terminals AB , Fig. 11.1, parallel to a grounded conductor. If equal alternating voltages of opposite phase are applied to the terminals, the momentary excess positive and negative charges along the conductor are shown in Fig. 11.1a at an instant when A is positive. Charges of opposite sign are induced along the parallel conductor, as in a capacitor,

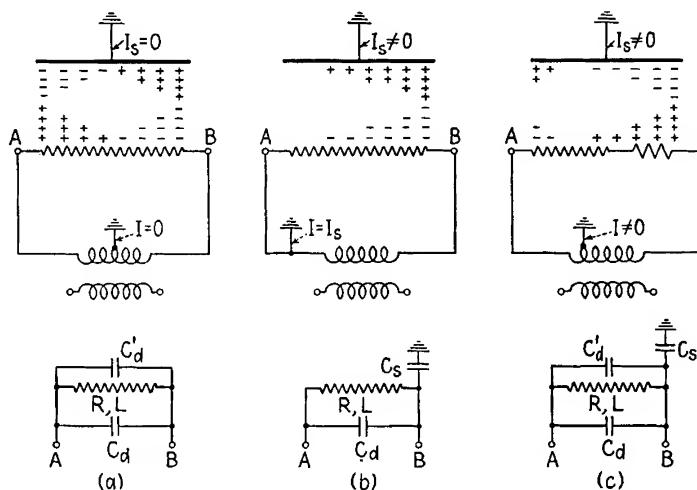


FIG. 11.1.—Representation of the electrical effects of an adjacent grounded conductor upon a resistor or coil: (a) when electrical balance exists; (b) when one end of the resistor or coil is grounded; (c) when the resistor or coil is unbalanced to ground.

but owing to the limited areas and large separation they are much smaller than the inducing charges. One-half period later the signs of the charges are reversed, but their magnitude and distribution are the same. There is thus produced an alternating current along the grounded conductor, but as the induced positive and negative charges are equal in magnitude no net current flows to ground. If the impedance of the grounded conductor is negligible, so far as measurements made at the terminals AB would determine, the effect of the adjacent conductor could be explained by connecting equal capacitances from A to ground and from B to ground. As no net current flows to ground, this is equivalent to adding to the representative circuit of the isolated element an additional parallel capacitor C_d' , the modified circuit of Fig. 11.1a being thus obtained.

Such a system is said to be balanced electrically with respect to ground, and the effect of the adjacent grounded conductor is to increase the shunting capacitance across the coil.

If the terminal A of the circuit element is grounded, Fig. 11.1*b*, the induced charge is of one sign, and the capacitive current is zero at A and a maximum at B . The current through B is the vector sum of that through A plus the unbalanced capacitive current to ground. The effect of the neighboring grounded conductor is to add to the representative circuit for the isolated element a capacitance to ground at the ungrounded end as shown below in Fig. 11.1*b*. If the geometry and resistance of the circuit element are not symmetrical about its geometrical center or if unequal alternating voltages are applied to A and B , Fig. 11.1*c*, unequal capacitive currents to ground flow through the terminals. In general, this is represented by placing unequal capacitances to ground at both terminals. In the special case of Fig. 11.1*c*, where a portion of the element remains at ground potential, the capacitive currents to ground partly cancel in the grounded conductor and the effect is to add a shunt capacitance to the representative circuit of the isolated element and a capacitance to ground at the terminal carrying the larger capacitive current.

The capacitive coupling between the surroundings and the unshielded inductance or resistance coil of Fig. 11.2*a* makes its impedance depend upon the geometry of neighboring objects and the nature of its potential variations relative to them. The coupling also permits undesired voltages and currents to be induced either in adjacent conductors owing to the action of the coil or in the coil owing to external potential variations. If a conducting shield encloses the element, Fig. 11.2*a*, and is connected to one terminal, the distributed capacitance of the coil is increased by an amount C_d' but the impedance between the terminals is made definite. The shield also prevents any interaction between the potential gradient along the coil and external conductors. Any capacitive current to the surroundings is confined to the terminal to which the shield is attached. The capacitive coupling between this terminal and the surroundings is still dependent upon their relative positions and potentials, but it can be rendered definite and undesired capacitive currents to other elements prevented by placing a grounded shield around it, Fig. 11.2*c*. Both the impedance of the element and its capacitance to ground now are made definite. Multiple shields of this sort are frequently employed in bridge

circuits. For many purposes the grounded shield of Fig. 11.2*c* may be used with the inner shield removed. This single shield prevents capacitive coupling between the coil and neighboring circuits and makes its impedance independent of changes in the geometry of the surroundings. Its capacitance to ground, however, depends upon the potential variations of its terminals with respect to ground.

Coupling due to mutual inductance is most important for two inductors but may occur between any two elements. It can be minimized by separating the elements or orientating them so as to produce minimum flux linkage. Induced voltages in a pair of leads

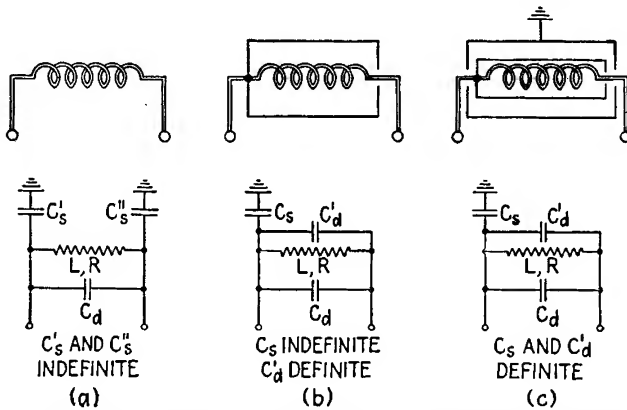


FIG. 11.2.—The effect of a shield in controlling the capacitance to adjacent objects and in fixing the impedance of an element.

may be minimized by keeping the two wires close together in the form of a twisted pair.

Magnetic shielding may be accomplished by enclosing the coil in a thick container of large permeability or small resistivity. The former is more effective at low frequencies where the shield provides a closed magnetic circuit for the flux, thus reducing the external magnetic field to a minimum. The high-permeability shield also causes an increase in the self-inductance of the coil, an increase in its resistance due to eddy-current and hysteresis losses, and an increase in its capacitance C_d . The usefulness of such a shield is reduced at high frequencies owing to a decrease in the effective permeability and to increased losses. Under these circumstances the shield should be made of a metal of high conductivity such as aluminum or copper. The shielding is due in this case to the induced currents in the surrounding metal whose magnetic field, by Lenz's law, opposes the magnetic field of the coil and thus tends to

cancel it at external points. The effectiveness of such a shield varies directly as the square root of the conductivity and of the frequency.

The thicker the shield, the better its shielding; for the strength of the magnetic field decreases exponentially with the thickness of metal penetrated. The shielding action will be lessened by any joint that introduces resistance in the path of the eddy currents. The magnetic field due to the currents in the shield likewise reduces the field of the coil and therefore slightly lowers its self-inductance. The energy loss due to the shield currents increases the series resistance of the coil, and the presence of the shield around the coil increases the value of C_d .

It is possible to construct a shield that will decrease the capacitive coupling without appreciably affecting the coupling due to mutual inductance. By slotting the conducting shield the eddy currents that produce the magnetic shielding may be greatly reduced without much change in the electric shielding. Such shields are sometimes employed between the windings of a transformer to prevent electric coupling between the input and output circuits.

12. Relation of Electrical Measurements to the Circuit Representation of an Element.—It has been shown possible to predict the electrical behavior of an actual inductance coil over a range of frequency extending upward from zero on the basis of an equivalent parallel combination of a fixed L , R , and C_d . As the frequency is increased, the skin effect and proximity effect require the value of R in this circuit to be progressively increased. Other losses due to imperfect dielectrics and eddy currents in adjacent conductors also affect R . Accompanying the change in R are smaller changes in L and still smaller changes in C_d .

The complexity of the behavior of such a coil in a practical circuit emphasizes the difficulty of predicting its performance at high frequencies on the basis of its low-frequency characteristics. For high frequencies the representative circuit is useful mainly as guide to the general behavior of the element. Numerical calculations based on it have validity only over a limited range of frequencies, and the values of L , R , C_d used must be those determined for that range.

These values are determined in general, not by extrapolation from the low-frequency case, but by measurements on the element made in the frequency range of interest. These measurements

involve the determination of the relation between the current through the element and the voltage across it and yield actual values of Z_a or Y_a that are definite at any given frequency. They may be described in either the polar or the rectangular form. In the latter form, Z_a leads directly to a description of the equivalent series circuit in terms of R_a and X_a ; or of R_a and L_a , where L_a is the value of the actual series, lumped inductance which would account for the reactive effects. In general, it is these equivalent series characteristics which are meant when the resistance R and the inductance L of an r-f coil are specified. The series resistance R_a accounts for all the energy lost, and $|I|^2 R_a$ gives the total power dissipation. For this reason it is sometimes referred to as the effective resistance of the element.

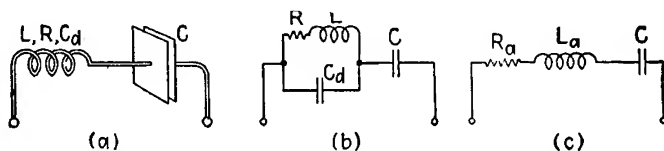


FIG. 12.1.—Series circuit of an actual inductor and capacitor, and its representative circuit.

For many purposes these values of R_a and L_a suffice to describe the behavior of the coil over a moderate range of frequencies where their changes in value are negligible. In particular, the series circuit containing an actual coil and capacitor, Fig. 12.1a, with the representative circuit in (b), may be described adequately by the series circuit of (c), in which R_a and L_a are assumed constant over a moderate range of frequencies. Fortunately, the capacitance of a good capacitor is constant over a wide band of frequencies and although its losses vary they may be neglected in comparison with the losses due to R_a .

It was shown in Fig. 3.1 that for any coil there is a range of frequencies where $Q_a = (\omega L_a / R_a)$ is almost constant despite variations in L_a and R_a . The variation of impedance of the coil and capacitor combination, Fig. 12.1a, can be described more accurately in this frequency range in terms of the constant Q_a of the coil and the resonant frequency of the combination as determined by direct measurement. The expression for Z may be written

$$Z = R_a + j \left(\omega L_a - \frac{1}{\omega C} \right) \quad (12.1)$$

$$\frac{Z}{R_a} = 1 + j \left(\frac{\omega L_a}{R_a} - \frac{1}{\omega C R_a} \right) = 1 + j Q_a \left(1 - \frac{\omega_r^2 L_r}{\omega^2 L_a} \right) \quad (12.2)$$

where ω_r is the series-resonant frequency of the coil and capacitor combination, L_r is the value of L_a at the resonant frequency, and use has been made of the relations

$$R_a = \frac{\omega L_a}{Q_a} \quad \text{and} \quad \omega_r^2 L_r C = 1 \quad (12.3)$$

In (12.2) R_a and L_a are functions of frequency that can be evaluated by measurement. However, if $C \gg C_d$, as is usually the case in practice, and the frequency deviation from resonance is small, the value of L_a will not change greatly for a reasonable range of frequencies and L_r/L_a may be taken as unity. Although the variations of R_a and L_a differ greatly from coil to coil, the rate of increase of R_a is always greater than that of L_a . In some cases the constancy of $Q_a (= \omega L_a/R_a)$ is due mainly to a linear increase of R_a with frequency, while L_a is almost constant. Under these circumstances (12.2) may be written

$$Z = w R_r \left[1 + j Q_a \left(1 - \frac{1}{w^2} \right) \right] \quad (12.4)$$

and

$$|Z| = \frac{1}{|Y|} = w R_r \sqrt{1 + \tan^2 \theta_z}$$

where

$$\tan \theta_z = Q_a \left(1 - \frac{1}{w^2} \right) \quad (12.5)$$

Here $w = \omega/\omega_r$, and R_r is the measured value of the equivalent series resistance at the resonant frequency.

The behavior of an actual coil and capacitor thus is predicted under these conditions in terms of constants determined in the frequency range of interest. The shape of the resonance curve is only slightly different from that obtained for an ideal series circuit with fixed lumped constants, but the phase-angle variation is not quite so symmetrical. For circuits of moderate Q_a the differences are minor, and the same relations exist between Q_a and the half-power frequencies measured in terms of circuit response. Then Q_a may be taken as the Q determined by laboratory measurements based on band-width determinations.

CHAPTER IV

MEASUREMENT OF CIRCUIT ELEMENTS

The impedance of a circuit element can be measured by direct determination of the current and voltage relation at its terminals. The accuracy of such measurements is limited by the characteristics of the instruments used. In general, an instrument can compare two voltages or currents more accurately than it can indicate their actual values. Therefore methods of measurement frequently are employed that involve comparison of voltages or currents. A further increase in accuracy is obtained by so-called "null" methods of measurement, in which a device serves only to indicate that a current or voltage has been reduced to zero. Some such circuits establish a relationship among the magnitudes of L , C , R , and ω . Bridge measurements are special cases of such null methods.

1. Theory of A-c Bridge.—The Wheatstone bridge, which was developed for the comparison of d-c resistances, may be extended

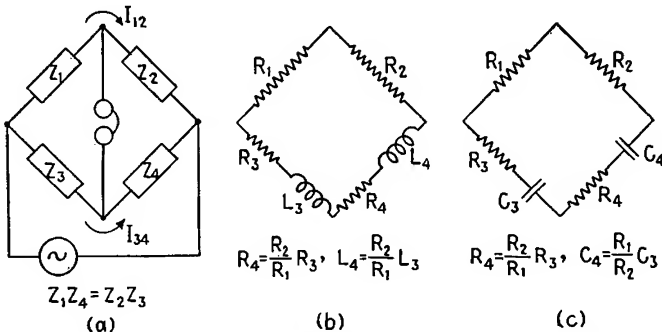


FIG. 1.1.—(a) Basic Wheatstone bridge circuit; (b) circuit for comparison of inductors; (c) circuit for comparison of capacitors.

to compare impedances. The impedances are connected as shown in Fig. 1.1, a source of alternating voltage is bridged across two of the opposite corners, and an a-c detector is bridged across the other two corners. If the impedances are adjusted so that no voltage exists across the detector, then no current flows through it, and the current through Z_1 equals that through Z_2 . Similarly, the

current through Z_3 is the same as that through Z_4 . The voltage drops across Z_1 and Z_3 are equal, as are those across Z_2 and Z_4 . Therefore,

$$Z_1 I_{12} = Z_3 I_{34} \quad \text{and} \quad Z_2 I_{12} = Z_4 I_{34} \quad (1.1)$$

where I_{12} and I_{34} are the currents through the upper and lower branches. Dividing one equality by the other,

$$\frac{Z_1}{Z_2} = \frac{Z_3}{Z_4} \quad \text{or} \quad \frac{|Z_1|/\theta_1}{|Z_2|/\theta_2} = \frac{|Z_3|/\theta_3}{|Z_4|/\theta_4} \quad (1.2)$$

For balance, the following relations exist between the magnitudes and angles of the impedances:

$$\frac{|Z_1|}{|Z_2|} = \frac{|Z_3|}{|Z_4|} \quad \text{and} \quad \theta_1 - \theta_2 = \theta_3 - \theta_4 \quad (1.3)$$

Cross multiplying the ratios of (1.2),

$$Z_1 Z_4 = Z_2 Z_3 \quad \text{or} \quad Y_1 Y_4 = Y_2 Y_3 \quad (1.4)$$

The products of the impedances or admittances of the opposite arms are equal. The two conditions of (1.3) arise from the fact that at balance both the magnitude and the phase of the sinusoidal voltage at the two detector terminals must be the same. In general, the balance requires the successive adjustment of at least two elements in the bridge, and it allows two unknown parameters to be evaluated in terms of other known standards. The impedances to be compared may be elements with distributed characteristics, but in general the bridge calculations are based on lumped values of R , L , and C .

In the following descriptions of several bridges, the source and detector are omitted, as their positions can be interchanged without effect upon the equations of balance.

2. Simple Types of Bridges.—A bridge circuit for the comparison of one inductor with another is shown in Fig. 1.1b. Applying (1.4), the general equation of balance,

$$R_1(R_4 + j\omega L_4) = R_2(R_3 + j\omega L_3) \quad (2.1)$$

$$R_1 R_4 + j\omega R_1 L_4 = R_2 R_3 + j\omega R_2 L_3 \quad (2.2)$$

Upon equating reals and imaginaries,

$$R_1 R_4 = R_2 R_3 \quad \text{and} \quad R_1 L_4 = R_2 L_3 \quad (2.3)$$

which are the conditions at balance; or

$$\frac{R_1}{R_2} = \frac{R_3}{R_4} \quad \text{and} \quad \frac{R_1}{R_2} = \frac{L_3}{L_4} \quad (2.4)$$

Since ω cancels, the balance is independent of the applied frequency.

The same type of circuit may be used for the comparison of capacitors, Fig. 1.1c. By similar calculation,

$$\frac{R_1}{R_2} = \frac{R_3}{R_4} \quad \text{and} \quad \frac{R_1}{R_2} = \frac{C_4}{C_3} \quad (2.5)$$

If R_4 and L_4 are unknown in Fig. 1.1b, a balance is obtained readily by varying the standard elements R_3 and L_3 . The accuracy

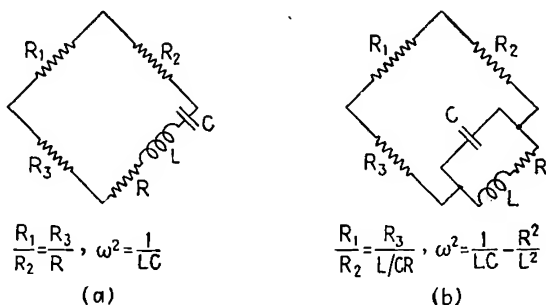


FIG. 2.1.—Resonance bridges.

is improved by making the ratio arms R_1, R_2 equal. If the standard of inductance available is a fixed inductor, the balance can be made by varying the ratio R_1/R_2 and R_3 ; but the equations (2.4) show that, as the ratio R_1/R_2 is adjusted, the balance conditions for both resistance and inductance are affected. Each successive change in one variable disturbs the balance condition for the other, causing the balance point to “crawl,” resulting in a slow approach to a final balance. The values of L_4 and R_4 obtained are the actual series equivalent resistance and inductance of the coil at the applied frequency. The values of R_3 and L_3 of the standard must also be those at the applied frequency.

In the two bridges of Fig. 2.1 the phase-angle condition of (1.3) requires that resonance exist in the fourth arm in order that it be a pure resistance at balance. The balance conditions can be calculated as before or can be based on the values of R and ω at series and parallel resonance.

$$\left. \begin{array}{l} \text{Series-resonance bridge } \frac{R_1}{R_2} = \frac{R_3}{R} \quad \text{and} \quad \omega^2 = \frac{1}{LC} \\ \text{Parallel-resonance bridge } \frac{R_1}{R_2} = \frac{R_3}{L/CR} \quad \text{and} \quad \omega^2 = \frac{1}{LC} - \frac{R^2}{L^2} \end{array} \right\} (2.6)$$

The frequency appears explicitly in the balance condition, and bridges of this sort may be used to measure frequency in terms of L , R , and C . If a resonance bridge is used to measure L and R , the frequency of the generator must be known. If the waveform of the generator is not sinusoidal, the harmonic frequencies will

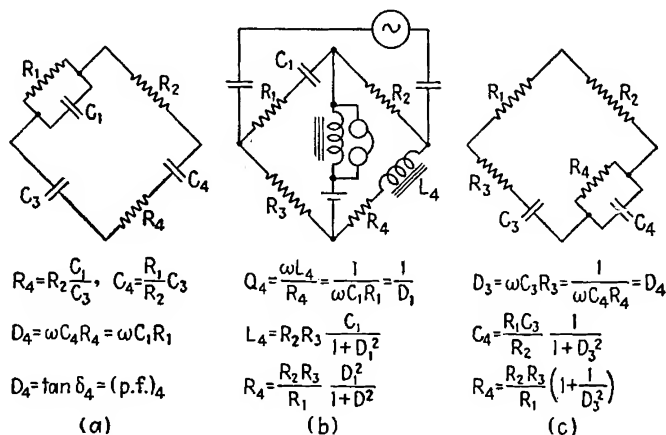


FIG. 2.2.—(a) Schering, (b) Hay, and (c) Wein bridge circuits and their conditions of balance.

still be heard in the detector when the bridge is balanced at the fundamental frequency. To obtain a balance under these conditions it is necessary to use a detector tuned to the fundamental or to make use of the frequency discrimination of the ear and reduce the fundamental note to zero against the harmonic background.

Resonance-bridge networks sometimes are used to suppress the fundamental so as to allow the residual harmonic content of a distorted wave to be measured.

Other simple bridges are shown in Fig. 2.2, together with their balance conditions. The Schering bridge is convenient for measuring the capacitance and dissipation factor of a capacitor. Its balance is independent of frequency, and successive adjustments are independent if R_1 and C_1 are varied. However, a fixed resistor can be made with smaller residuals than a variable one; and if R_1 is fixed, balance still can be obtained by varying the two air capaci-

tors C_1 and C_3 . This gives a smooth variation for balance; but the two adjustments are interdependent, and the balance crawls. The bridge is adaptable to measure a wide range of capacitances down to very low values. It may be used also to measure the high-voltage characteristics of the capacitor C_4 in arm 4. The high-voltage generator is connected between the top and bottom corners of Fig. 2.2a with the top terminal grounded. The variable elements are R_1 , R_2 , C_1 ; together with the detector, they remain at relatively low potentials if the impedance of the grounded half of the bridge is kept small in comparison with the half containing the unknown and the standard fixed low-loss capacitor C_3 .

The Hay bridge, Fig. 2.2b, may be conveniently used to measure the inductance and resistance of an iron-core choke since the connections of supply voltage and detector shown in the diagram allow a direct current to be maintained through the choke while its a-c characteristics are determined. For measurements on an air-core inductor the usual supply and detector circuits are sufficient. Under these conditions there is only one inductor in the bridge circuit, and magnetic coupling with other elements is reduced. The Hay bridge also allows an inductor to be measured in terms of a capacitor, which is in general a much more satisfactory standard at radio frequencies than an inductor. The r-f inductor is connected in place of the iron-core inductor indicated by R_4L_4 , Fig. 2.2b, and the quality factor of the coil is determined readily at the frequency of measurement from C_1 and R_1 at balance. (If the quality factor is low, more satisfactory results are obtained by placing R_1 in parallel with C_1 , forming a Maxwell bridge.) The balance point of the Hay bridge depends upon the frequency. In fact, the bridge may be used as an audio-frequency measuring device with a linear scale, since the equations of balance give

$$\omega = 2\pi f = (R_2R_3L_4C_1 - L_4^2)^{-\frac{1}{2}}R_4$$

By suitable choice of circuit constants the factor multiplying R_4 can be made equal to 2π and then the frequency in cycles/second equals R_4 in ohms. The bridge is balanced by varying R_1 and R_4 .

The Wein bridge of Fig. 2.2c is similar to the bridge of Fig. 1.1c. If C_4R_4 represent the unknown capacitor, these two bridges measure it alternatively in terms of either a parallel or a series combination of C and R . The Wein bridge balance does not crawl if R_3 and C_3 are varied. Its balance depends upon the frequency, and this property is employed in several applications. If

R_4 is known, it is possible to evaluate C_4 in terms of frequency and resistance without reference to C_3 .

3. Application of Bridge Methods.—Certain considerations are applicable, not only to the bridges described, but to bridge circuits in general. The method of obtaining the balance, the arrangement of the circuit elements, the nature of the standards used, all have their effect upon the speed and accuracy of the determination.

Balance is obtained in a bridge by adjusting one bridge parameter until a minimum response is produced in the detector and then changing a second parameter to reduce the response to a still lower value. These adjustments are made alternately until no further reduction in the output of the detector is observable. From the values of the known circuit parameters the impedance at the terminals of the unknown is calculable. The unknown impedance generally is represented in the bridge circuit and appears in the equations for balance as a series or parallel combination of resistive and reactive components. The corresponding values of L , C , and R calculated from the balance equations have an accuracy that depends upon many factors. In general, the accuracy of the bridge is improved if the magnitude of the impedance in each of the four arms is approximately the same. Although the equations of balance are not affected if the generator and detector are interchanged, the current in the detector for a given unbalance of the bridge will be greater the more nearly the internal impedance of the generator is matched to the input impedance of the bridge considered as a load on the generator. The detector also should be matched to the internal impedance of the equivalent generator at the detector terminals of the bridge. If the detector and generator impedances are not equal, the sensitivity of the bridge sometimes may be improved by reversing their positions. The sensitivity also may be improved by the use of proper matching transformers.

If an amplifier is used in connection with the detector, matching considerations are of limited importance in determining the range of the variable over which no perceptible change occurs in the detector indicator. The accuracy depends upon the sharpness of this null, which with an amplifier is apt to be limited by the presence of extraneous voltages due to undesired couplings with the surroundings or between the elements of the bridge. Also, when harmonics are present in the source they are not balanced out at the same setting as the fundamental if the bridge balance depends upon

the frequency or if the parameters of the bridge arms are functions of frequency. Undesired couplings are reduced by proper shielding, and troubles due to harmonics may be minimized by the choice of a suitable generator and detector.

The accuracy of any bridge measurement of course is limited by that of the standard used. It should be remembered that every standard has residual characteristics whose effects must be considered, particularly at high frequencies. It is possible to reduce these residuals more effectively in fixed standards than in variable ones. For accurate work at high frequencies a standard should have at least one shield in order to make its distributed capacitance definite. This shield frequently is permanently attached to one terminal of the element. A second shield surrounding the first, and insulated from it, is convenient as it allows the capacitance to ground of the element to be fixed.

Standards of inductance are useful mainly for low-frequency measurements because of their comparatively large residual parameters, which vary with frequency. In the audio-frequency range a variable mutual inductor, which may also serve as a variable self-inductor, is a convenient standard. Fixed resistors may be used over much wider ranges of frequency if precautions are taken to reduce the skin effect. In variable resistors, generally available as decade boxes, the varying capacitance and inductance of the switching mechanism are possible sources of error. By proper design their effect may be reduced, and certain "compensated" boxes arrange the switching so as to make the residual self-inductance independent of the positions of the switches. For certain types of measurement this reactance component can be canceled by proper adjustment, in which case it causes no error in the determination.

The variable air capacitor is the most usable standard at radio frequencies. The variation of capacitance is smooth and can be made linear over most of the scale. Although the resistance associated with an air capacitor varies in a complicated manner with frequency, its effects generally are negligible in comparison with other losses in the circuit. The capacitance is nearly independent of frequency up to radio frequencies, where the effect of the residual series inductance causes the capacitance to increase with frequency, at first slowly and then more rapidly.

Over a greatly reduced range of frequencies, fixed mica capacitors may be used. Paper capacitors have a still more limited use

as standards. The losses of such solid-dielectric capacitors are much larger than those of air capacitors, and as the applied frequency approaches zero the capacitance may increase slightly owing to dielectric absorption. In the radio-frequency range, specially designed transmission lines also may be used as standards.

The nature of the generator depends upon the frequency at which the measurements are made. Power-line voltage at 60 cps may be used, but special generators delivering higher frequency signals are generally employed. The microphone hummer, which consists of a tuning fork maintained in oscillation by an electro-mechanical system, can be made to give a reasonably pure signal of fixed frequency. Audio oscillators of either the beat-frequency or the resistance-capacitance type give a close approximation to a sinusoidal voltage, and their frequency is conveniently variable. In general, the purity of waveform is improved by restricting their operation to low output levels. At radio frequencies a modulated-signal generator is a convenient source.

The most commonly used detector at low frequencies is a pair of headphones, used with or without an amplifier. At 60 cps the ear is not as sensitive as at higher frequencies, and an a-c vibration galvanometer may be used. If this is tuned to 60 cps, it becomes a sensitive device and one that tends to ignore stray voltages and harmonics. The headphones also may be tuned to the bridge frequency, or tuned circuits may be inserted in the amplifier in order to achieve a sharper null due to the reduction of background noise. At radio frequencies a well-shielded all-wave radio receiver is a convenient detector.

In practice, the simple conditions of the ideal bridge network are seldom achieved, particularly at high frequencies. The inaccuracies caused thereby may be partly avoided by using the bridge to compare directly the unknown and a variable standard. The bridge is balanced with the unknown connected in one arm. The unknown is then replaced by the standard and the balance restored by varying the standard. This is known as the "substitution" method. This simple method is useful only if the residuals of the unknown and the standard are negligible. It is employed mainly for the comparison of capacitors, particularly small air capacitors. If residuals exist, the balance frequently may be restored by minor adjustments in the bridge itself and much of the advantage of the method retained.

Another method that reduces the errors due to the bridge is to

make a preliminary balance on a four-arm bridge and then to connect the unknown to one arm of the bridge. The unknown is connected in series if it has a low impedance, in parallel if it has a high impedance, and the bridge is then rebalanced. The impedance of the unknown element is calculated from the difference of the two bridge settings. This may be described as the "difference" method. Although it is not as independent of possible bridge errors as the substitution method, it may reduce their effect appreciably.

4. Precautions at High Frequencies.—When the generator, standards, unknown impedance, and detector are connected together according to the wiring diagram of one of the bridge circuits, there are at the same time other inductive and capacitive couplings (not indicated on the diagram) between each of the elements and their surroundings. A current in one branch may induce a voltage in another branch, or a charging current may flow in a branch because of its capacitance with respect to another branch or to ground. These undesired voltages and currents increase with the frequency, and the result is that at sufficiently high frequencies the actual bridge circuit is far from the ideal one of the diagram. The effect of these couplings is to make the impedance of the bridge arms depend upon the surroundings, to decrease the sharpness of the null at balance, and to give an incorrect value for the impedance of the unknown.

The inductive coupling is greatest between coils. It can be reduced by using coils having small external fields such as toroids or other coils with closed magnetic circuits or by separating and orienting coils approximately at right angles so that their magnetic fields produce a minimum of interaction. Matching transformers, when used for the source and detector, should be well separated. If the number of coils can be reduced to one by proper choice of a bridge circuit, the magnetic linkage is mainly with the loops formed by the wires themselves. This can be reduced by keeping the loops small or by running the connecting wires close together in twisted pair or coaxial line. However, these precautions will increase the stray capacitance.

The capacitive coupling *between* elements can be reduced by surrounding each element and its leads by a grounded shield as shown in Fig. 4.1a. The generator G and the detector D are each represented by one of the windings of the matching transformers. The impedances Z_1, Z_2, Z_3, Z_4 , as well as the generator and detector, are shown surrounded by shields, insulated from them and in

practice extending from terminal to terminal of the bridge. If the corresponding shields $S_1, S_2, S_3, S_4, S_a, S_D$ are connected to ground, but no part of the bridge is grounded, the equivalent circuit is that of Fig. 4.1*b*, where the four capacitances from the bridge terminals a, b, c, d to ground are due to the stray capacitance of the shields on the bridge arms and also the generator and detector shields. Although the shields increase the capacitance of the elements to ground, they make that capacitance definite, and no extraneous voltages are produced in the bridge arms by capacitive coupling between arms. The four capacitances provide paths between

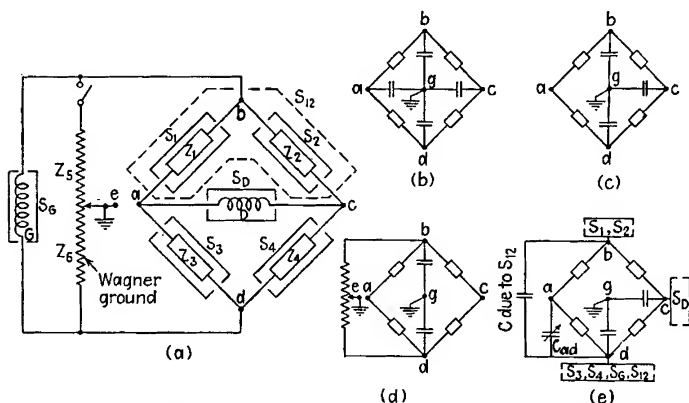


FIG. 4.1.—(a) Bridge with shielded impedances and a Wagner ground; (b) Wagner ground disconnected and all shields grounded; (c) terminal a grounded also; (d) Wagner ground connected and adjusted; (e) Wagner ground removed, S_{12} added and connected to terminal d , and other shields connected as shown.

opposite corners that shunt the generator and detector. These shunt paths alone may reduce the sensitivity slightly but cannot affect the balance. In addition, owing to the common ground, they shunt the arms of the bridge and may cause an error in its indication if the impedances in the arms are sufficiently high. If with the shields all grounded the terminal a is grounded as well, there will be no potential difference between a and g , therefore no current, and the capacitor C_{aa} may be omitted as in Fig. 4.1*c*. The arms bc and cd are still shunted by stray capacitances to ground.

The "Wagner ground" may be used to reduce still further the effect of stray capacitances to ground. It is shown in Fig. 4.1*a* as an added two-element potential divider, Z_5, Z_6 . The impedance of this branch must be small in comparison with that of the stray capacitances. At low frequencies it may be a wire-wound poten-

tiometer, and for high frequencies it often consists of a pair of variable air capacitors. In using the Wagner ground the first step is to obtain a rough balance on the bridge. If the effect of stray capacitances is serious, this null is apt to be poorly defined. The detector connection at c is broken, the detector is connected between a and e , and a balance is obtained by varying Z_5 and Z_6 . Upon returning the detector to its original position between a and c and rebalancing, a much sharper null is obtained. This process is repeated until no further improvement results in the balance. Generally only one or two switching operations are necessary. The Wagner ground balance adjusts the potential of terminal a to ground. The bridge balance ensures that the potential of terminal c is the same as that of a , and therefore both a and c remain at ground potential even though not directly connected to ground. The equivalent circuit reduces to that of Fig. 4.1*d*, where the capacitances C_{ag} and C_{cg} are neglected. The only remaining capacitances simply shunt the generator and have no effect on the bridge balance. The Wagner ground therefore not only ensures sharper nulls but eliminates the error due to the ground capacitances. It may be employed with profit in many bridge circuits, even though the elements are not separately shielded, by connecting it to a shielded container surrounding the entire bridge. Some advantage is gained by its use even if no shield at all is used and the Wagner ground is connected to any good ground in the neighborhood. If the ab and ad arms of the bridge are the fixed-ratio arms, very few adjustments of the Wagner ground are necessary.

Control of the capacitance to ground can be obtained also by connecting the shields to various terminals of the bridge. If the shields S_1, S_2 are connected to terminal b ; S_3, S_4, S_6 to terminal d ; and S_5 to terminal c , the equivalent circuit is again that of Fig. 4.1*c*, although no parts of the bridge and none of the shields are grounded. The capacitance C_{bg} is now the capacitance of S_1 and S_2 to ground, that of C_{cg} is the capacitance of the detector shield to ground, and C_{dg} is the sum of the capacitances to ground of shields S_3, S_4, S_6 . If a second shield S_{12} is added surrounding shields S_1 and S_2 (as indicated in Fig. 4.1*a* by a dotted line) and connected to terminal d , the effect is to change the equivalent circuit to that of Fig. 4.1*e*. The capacitance to ground of terminal b is reduced to zero, and there is added instead a shunt capacitance from b to d that cannot affect the bridge balance. The arm cd is still shunted by the series combination of C_{cg} and C_{dg} . If now the terminal d is grounded,

only the capacitance C_{cg} due to the detector screen shunts the cd arm. The capacitance across the cd arm can be balanced by the small variable air capacitor C_{ad} , Fig. 4.1e. After this preliminary balance an unknown can be placed in parallel with the cd arm, the balance restored by varying the ad arm, and the method of differences used. Alternatively, both the arms cd and ad can be replaced by other elements having their shields connected to d and the effects of stray capacitances to ground remain eliminated provided that the ratio arms ab and bc are not altered.

A further advantage is obtained if equal-ratio arms ab and bc are used. If the shield on the detector coil is disconnected from c and is grounded at the center, it adds equal capacitances from a to g and from c to g if precautions are taken to make it symmetrical. When the bridge is balanced, the circuit is symmetrical about the vertical line bd and the capacitive currents through C_{ag} and C_{cg} cannot affect the balance. This assumes that the shields S_1 and S_2 likewise have equal capacitances to ground.

Thus it is possible by the use of shields to decrease the capacitances between the arms of the bridge and to render their impedances definite. By suitable connection of the shields the capacitances to ground may be thrown across the source or the detector or across a low-impedance arm. They may likewise be balanced out by a preliminary adjustment or by resorting to symmetrical construction. It is only by the use of such carefully designed shields that bridge methods have been extended successfully to radio-frequency measurements.

5. Bridged T and Parallel-T Measurements.—Another null method of measurement employs a three-terminal network, as opposed to the four terminals of a bridge circuit. The bridged T, Fig. 5.1a, consists of a T-shaped arrangement of impedances Z_1 , Z_2 , Z_3 , connected between the terminals d , g , and the common terminal e , h . Across the top of the T, from terminal d to terminal g , is bridged an impedance Z_4 . If a generator is connected across the terminals d , e , it is possible, by suitable adjustments, to make the output voltage at g , h approach zero.

The conditions necessary for this null in output voltage may be determined by making use of the T to Π transformation of Chap. V, Sec. 4. It is shown there that the T connections between the three terminals enclosed in the dashed rectangle of Fig. 5.1a may be replaced by an equivalent circuit consisting of Z_A , Z_B , Z_C , arranged in a Π connection between the same terminals, Fig. 5.1b. The

resultant impedances Z_A , Z_B , Z_C may not be physically realizable, but any calculations based upon the equivalent circuit lead to exactly the same currents and voltages at the three terminals as for the original circuit. If the impedance of the parallel combination Z_A , Z_B is infinite, it is apparent from Fig. 5.1b that there can be no current through the detector.

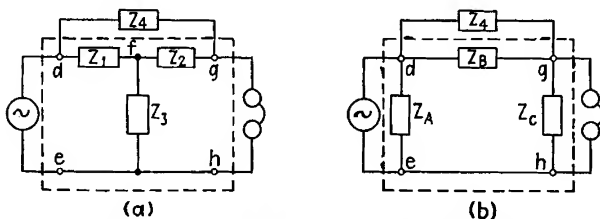
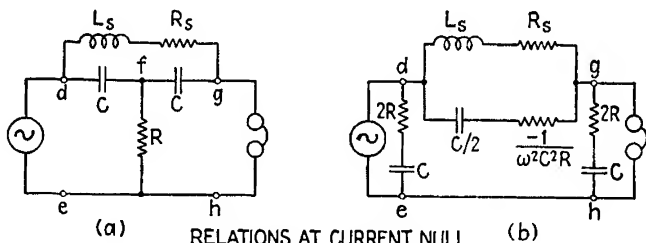


FIG. 5.1.—(a) Bridged T; (b) equivalent circuit.

As an example, consider the inductance coil, represented in Fig. 5.2a by the series combination of L_s and R_s , bridged across the capacitive arms of a T. If the T is transformed to a Π by use of (4.10) to (4.12), Chap. V, the equivalent circuit of Fig. 5.2b is obtained. The horizontal section of the Π between terminals d and g is equivalent to a capacitance of magnitude $C/2$ and a *negative* resistance of magnitude $1/\omega^2 C^2 R$. If the parallel combination



RELATIONS AT CURRENT NULL

$$L_s = \frac{2}{\omega^2 C}, \quad R_s = \frac{1}{\omega^2 C^2 R}, \quad Q = \omega CR$$

FIG. 5.2.—(a) Bridged T; (b) equivalent circuit.

between terminals d and g is adjusted so that both reactance and resistance of one branch are equal and opposite in sign to those in the other branch, there results a parallel-resonant circuit with zero losses whose impedance is infinite. The conditions for a voltage null at g , h are therefore

$$\omega L_s = \frac{2}{\omega C} \quad \text{and} \quad R_s = \frac{1}{\omega^2 C^2 R}$$

and the quality factor of the coil is equal to ωCR . Balance is obtained by varying R and each C , Fig. 5.2a. As C enters both balance conditions, the balance crawls.

This method has the advantage that the generator and detector have one side in common, which may be grounded. Care must be taken, however, to reduce the stray capacitance shunting R and to avoid stray coupling between terminals d and g .

Another form of the bridged T allows one side of the unknown coil to be grounded as well. In Fig. 5.3a the unknown coil is represented by the parallel combination of L_p and R_p . If the T in this circuit is transformed to a Π , the circuit of Fig. 5.3b is obtained, where R is bridged across the equivalent series LCR combination, which forms the horizontal part of the Π . The values

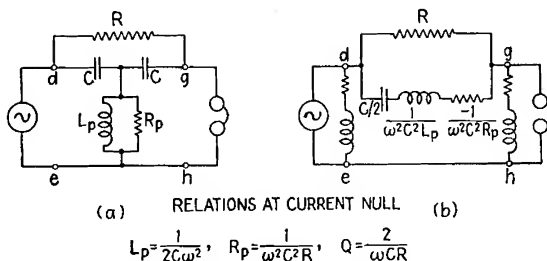


FIG. 5.3 — (a) Bridged T; (b) equivalent circuit.

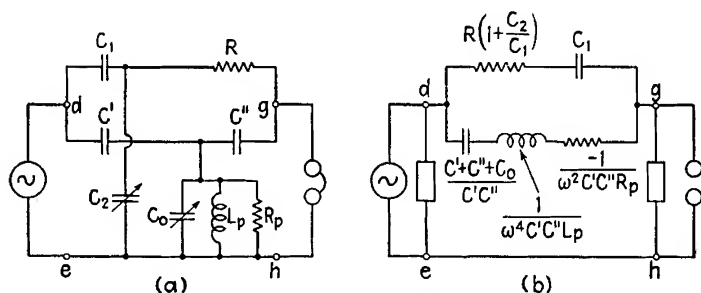
of the series components are shown on the diagram. If the series inductance and capacitance have equal and opposite reactances the d, g connection reduces to a parallel combination of a positive and negative resistance. If these are equal in magnitude, the impedance of the combination is infinite. To satisfy these conditions the null equations shown in Fig. 5.3 suffice.

The parallel, or twin-T, network consists of two T sections connected in parallel between the same three terminals. The circuit of Fig. 5.4 illustrates one possible combination, where the unknown coil is represented by the equivalent parallel combination of L_p and R_p . Upon transformation of the two T sections to equivalent Π sections, two parallel branches between d and g are obtained, Fig. 5.4b.

The equations of zero voltage transfer (*i.e.*, infinite impedance between d and g) given in Fig. 5.4 are satisfied by variation of C_0 and C_2 . Two important advantages follow from this, the balance conditions for L_p and R_p are now independent, and the resistor R is fixed and therefore can be made with smaller residual inductance

and capacitance. In addition, one side of both the standard capacitors C_0 and C_2 may be grounded as well as the generator and detector. This simplifies the shielding problem and allows the network to be used at relatively high radio frequencies.

The effects of stray couplings may be further reduced by applying the method of differences. The bridge is first adjusted to a null condition with a coil in place. Then the coil to be measured is connected in parallel with the first coil. The inductance L_p and the



EQUATIONS AT CURRENT NULL

$$L_p = \frac{1}{\omega^2 [C_0 + C'C'' (\frac{1}{C_1} + \frac{1}{C''} + \frac{1}{C_1})]}, \quad R_p = \frac{1}{\omega^2 RC'C'' (1 + \frac{C_2}{C_1})}$$

FIG. 5.4.—(a) Parallel, or twin-T; (b) equivalent circuit.

parallel resistance R_p of the second coil is easily expressed in terms of the difference in C_0 and C_2 for the two balances. If ΔC_0 and ΔC_2 represent the increases in C_0 and C_2 , then, for the second coil,

$$L_p = \frac{1}{\omega^2 \Delta C_0} \quad \text{and} \quad R_p = \frac{C_1}{\omega^2 RC'C'' \Delta C_2}$$

Either inductive or capacitive impedances may be measured in this way. For an unknown with a capacitive reactance, the value for ΔC_0 would be negative.

6. Single-circuit Resonance Methods.—Measurements based on the relations in a series loop at or near resonance do not have the inherent accuracy of null methods. However, the simplicity of the circuit makes it possible to obtain moderate accuracy at radio frequencies without the elaborate precautions against coupling required in more complicated networks.

Q Meter.—The Q meter uses the resonant rise of voltage in a series loop to determine the Q of the circuit and therefore the Q of the coil at the resonant frequency, if it can be assumed that the coil is responsible for the losses. In the basic circuit of Fig. 6.1a

an oscillator maintains a constant driving current I_d through the resistance R whose magnitude is small in comparison with R_s . This current is adjusted to produce a standard voltage across R , which then acts as low-resistance generator for the circuit. The low-loss air capacitor C is varied until E_c , as measured by a vacuum-tube voltmeter, is a maximum. For a low-loss circuit this corresponds to resonance, and L_s can be evaluated in terms of the frequency ω and the capacitance C . At the same time the voltmeter reading E_c as compared with the known applied voltage is a measure of Q . The scale of the voltmeter may be calibrated to read Q directly. Knowledge of Q , ω , and L_s enables R_s to be calculated. Such a meter allows a rapid comparison of the Q of various coils with an accuracy that is satisfactory for all but very low-loss or very high-loss inductors.

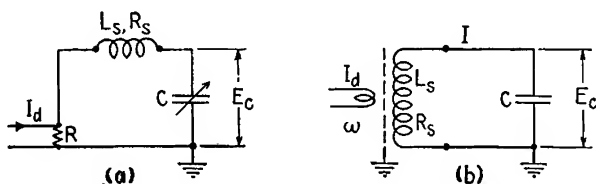


FIG. 6.1.—Resonant measurement circuits.

Resistance-variation Method.—The resistance-variation method makes use of the simple current-voltage relations in a series loop at resonance to determine the circuit resistance, which can often be assumed to be that of the coil. The series loop of Fig. 6.1*b* is excited by the voltage induced in the inductor due to the driving current I_d in an adjacent coil. A grounded shield is placed between the coil and the inductor to eliminate capacitive coupling between them, and the inductive coupling is made small enough to ensure that changing currents in the resonant circuit have no effect upon I_d ; therefore the induced voltage is constant. An r-f milliammeter, generally of the thermocouple type, indicates the current I .

The circuit is tuned to resonance and the magnitude of current I is measured. An additional resistor R_{added} is inserted in series with the loop and the circuit again tuned to resonance if necessary because of the residuals of the resistor. If the magnitude of I were reduced to half its original value, the original circuit resistance clearly would be equal to the added resistance R_{added} . Generally, however, several values of R_{added} are inserted, and the reciprocals of the corresponding currents are plotted against the values of R_{added} .

Fig. 6.2a. If the line obtained is extended to zero on the vertical scale (corresponding to infinite current), the intercept to the left of the axis gives the resistance that should be subtracted to make the total circuit resistance zero. This is of course equal to R_s if other losses can be neglected.

Reactance-variation Method.—If C is the only variable in Fig. 6.1b, a knowledge of the corresponding variation of the current is sufficient to evaluate the Q of the circuit. The quality factor and the value of L_s and R_s for such a circuit with variable capacitance were shown in Chap. II, Sec. 12, to be determined by the

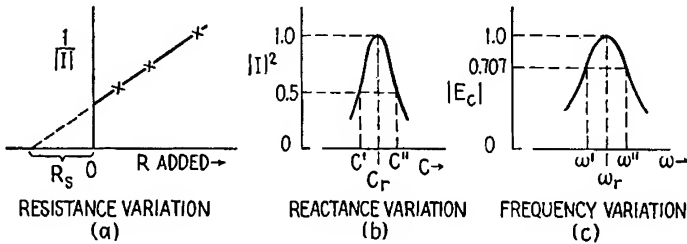


FIG. 6.2.—Evaluation of circuit parameters.

half-power value of C and the frequency. From (12.7) and (12.8), Chap. II, the equivalent series inductance L_s , series resistance R_s , and Q , are given by

$$Q = \frac{C'' + C'}{C'' - C'} \quad L_s = \frac{C'' + C'}{2\omega^2 C' C''} \quad \text{and} \quad R_s = \frac{C'' - C'}{2\omega C' C''} \quad (6.1)$$

The half-power values of C may be determined from a curve similar to Fig. 6.2b. Here the square of the current is plotted, since the response of a thermocouple meter commonly is proportional to $|I|^2$. In this case the half-power values of C are those which reduce $|I|^2$ to one-half its value at resonance.

Frequency-variation Method.—If the frequency of the constant-magnitude driving current I_d is altered and the variation of E_c is determined, it is possible to evaluate the circuit parameters in Fig. 6.1b. The voltage across C is given by the product of the current and the capacitor impedance; in magnitude values,

$$|E_c| = \left| \frac{\omega M I_d}{Z} \right|, \quad \left| \frac{1}{\omega C} \right| = \left| \frac{M I_d}{C} Y \right|$$

Therefore the variation of $|E_c|$, except for a constant factor, is of the same form for this circuit as the variation with frequency of the

admittance $|Y|$ of the series loop. It is possible to determine the half-power and resonant frequencies of the circuit from a curve similar to that of Fig. 6.2c. The half-power frequencies are those at which E_c drops to 0.707 of its value at the resonant frequency ω_r . From (11.10) and (11.5), Chap. II,

$$Q = \frac{\omega_r}{\omega'' - \omega'} \quad L_s = \frac{1}{\omega_r^2 C} \quad \text{and} \quad R_s = \frac{1}{Q_r \omega_r C} \quad (6.2)$$

Although these methods have been applied to the measurement of coils, it is possible to extend them all to the measurement of resistor characteristics. After a preliminary measurement of a coil by one of the described methods, an unknown resistor may be placed in series or in parallel with the coil (depending upon whether its resistance is small or large) and a new balance obtained. From the difference in the two measurements it is possible to calculate the characteristics of the added resistor.

7. Evaluation of Equivalent Circuit Constants from Measurements.—In the methods of measurement described, the constants determined for a coil were its equivalent series inductance L_s and series resistance R_s at one frequency. By determining the constants of the coil at several frequencies it is possible to determine the L , R , and C_d of the equivalent circuit, Chap. III, Fig. 2.1c, which was earlier shown to describe it.

Under the assumption that L , R , C_d are constant and of magnitudes suitable to describe a well-made coil, it is shown in Chap. III that, for all frequencies up to those near the self-resonant frequency of the coil, the equivalent series inductance L_s is related to L as follows,

$$\frac{L_s}{L} = \frac{1}{1 - w^2} \quad \text{where} \quad w = \frac{\omega}{\omega_{sr}} = \omega \sqrt{LC_d} \quad (7.1)$$

from (3.4), Chap. III. Here, ω_{sr} is the series-resonant frequency of the loop formed by L and C_d . Although in practice the resistance of the coil increases greatly with frequency owing to skin effect, the inductance L is more nearly constant, decreasing slightly. It is possible to identify L in the equivalent circuit with the low-frequency self-inductance of the coil L_{dc} and from (7.1) to obtain

$$\frac{L_s}{L_{dc}} = \frac{1}{1 - \omega^2 L_{dc} C_d} \quad (7.2)$$

Inverting, and dividing by L_{dc} ,

$$\frac{1}{L_s} = \frac{1}{L_{dc}} - \omega^2 C_d \quad (7.3)$$

L_s is the measured equivalent series inductance of the coil at the frequency ω . If $1/L_s$ is plotted against ω^2 , the linear variation of Fig. 7.1a is obtained if the inductance of the coil is constant. The value of L_{dc} is obtained from the intercept on the vertical axis, and C_d can be determined from the slope of the line.

For some coils the variation of the inductance L due to skin effect is sufficient to cause appreciable deviation from the straight

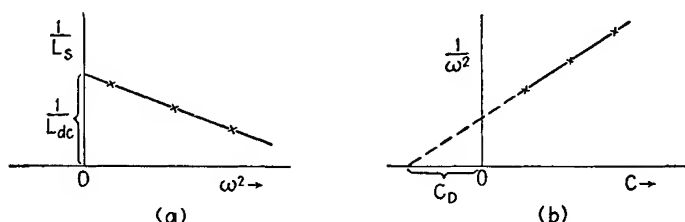


FIG. 7.1.—Evaluation of equivalent circuit constants.

line, corresponding to departure of L from its low-frequency value L_{dc} .

If the values of L_s and R_s are obtained by resonant-circuit methods, it is possible to evaluate C_d from the relation between the capacitance C and the frequency ω at which resonance exists. At resonance, ωL_s is equal to $1/\omega C$, and (7.3) may be transformed to

$$C + C_d = \frac{1}{L_{dc}} \frac{1}{\omega^2} \quad (7.4)$$

In Fig. 7.1b, $1/\omega^2$ is plotted against C , the resonant value of the tuning capacitor, at various applied frequencies. The intercept on the horizontal axis gives the value of C_d . The slope of the line depends on the value of the self-inductance. If it varies from its low-frequency value over the range of frequencies applied, the line becomes curved and the value of the C_d determined becomes a function of the frequency range.

CHAPTER V

NETWORKS AND IMPEDANCE MATCHING

1. Introduction.—A network may be defined as any combination of generators and impedance elements. A network composed of impedance elements but with no sources of electromotive force is a passive network, while one containing generators is an active network. There are two general objectives in the study of networks, first to express the impedance presented by the network between any two points, and second to express the current in or the voltage across any element of the network when a voltage is applied in any arbitrary manner. For the purposes at hand the study of networks will be confined to the relatively simple cases of two-, three-, and four-terminal networks.

Certain limitations on the scope of the network equations and theorems are imposed by the need for simplicity. First, all circuit elements of which the network and its external circuits are composed are assumed to be linear elements. Thus the values of the resistance, inductance, and capacitance of the various elements are assumed to be independent of the current or voltage; they may be and in general are functions of applied frequency. Mathematically speaking, the term linear means that the differential equations of the circuits are linear differential equations with constant coefficients. The general form of Ohm's law is the mathematical expression of the steady-state voltage-current ratio in linear circuits. If the applied voltage is doubled, the corresponding current is doubled, etc. In general, the linearity of all circuit elements is restricted by certain current and voltage limitations. Extreme currents cause heating and changes in dimension; high voltages may cause corona or a breakdown of insulation. However, by limiting the maximum voltages and currents it is possible to consider most ordinary resistances, inductances, and capacitances as linear elements. Many devices such as vacuum tubes, electrolytic capacitors, and iron-core inductances exhibit linear properties over limited ranges of their current-voltage characteristics. In these ranges they may be regarded as linear elements in the calculation of variational components of current.

If an alternating voltage of sinusoidal waveform at any given frequency is applied to a linear circuit element, the resulting current is of the same frequency and waveform and contains no harmonics (Chap. IX). On the other hand, the current flowing through a nonlinear circuit element for a sinusoidal applied voltage does contain harmonics. Which harmonics appear and whether or not the fundamental appears depend on the circuit and the nature of the nonlinearity.

2. Network Notation and Equations.—Since a network may consist of a large number of circuit elements in any arbitrary arrangement, it is desirable to have a system of notation by which the various quantities to be used in the network equations may be

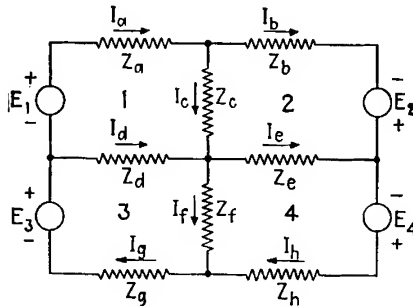


FIG. 2.1.—Example of a four-mesh network, branch currents indicated.

designated. Figure 2.1 shows a “four-mesh” network having a generator and impedances in each loop, or “mesh,” of the network. In Fig. 2.1 and in succeeding diagrams the symbol for current through each branch or around each loop is accompanied by an arrow, which indicates the *assumed positive direction*. This arrow does not indicate the actual direction of current flow since the current under consideration is the root-mean-square value I of the instantaneous current $i = \hat{I} \sin(\omega t + \theta)$ that flows for an equal time in each direction. Another caution about current directions: The arrows do not indicate that the currents in the several branches flow simultaneously in the indicated directions. They may have any phase relationship whatsoever, depending on the assumed positive directions and on the nature of the circuits. The assumed positive directions of externally applied voltages are indicated by plus and minus signs on the symbol for the voltage.

A branch is a path from one junction to another. In the network of Fig. 2.1, the branch current I_a flows through the impedance

Z_a , I_b through Z_b , etc. The Kirchhoff voltage equations for each closed loop of the circuit are

$$E_1 - I_a Z_a - I_c Z_c + I_d Z_d = 0 \quad (2.1)$$

$$E_2 + I_c Z_c + I_e Z_e - I_b Z_b = 0 \quad (2.2)$$

$$E_3 - I_d Z_d - I_f Z_f - I_g Z_g = 0 \quad (2.3)$$

$$E_4 - I_h Z_h + I_f Z_f - I_e Z_e = 0 \quad (2.4)$$

Equating junction currents to zero,

$$I_c = I_a - I_b \quad (2.5) \quad I_f = -I_h + I_g \quad (2.7)$$

$$I_d = -I_a + I_g \quad (2.6) \quad I_e = -I_b + I_h \quad (2.8)$$

Substituting these values in the voltage equations and regrouping terms,

$$E_1 - I_a(Z_a + Z_c + Z_d) + I_b Z_c + I_g Z_d = 0 \quad (2.9)$$

$$E_2 - I_b(Z_b + Z_e + Z_c) + I_h Z_e + I_a Z_c = 0 \quad (2.10)$$

$$E_3 - I_g(Z_d + Z_f + Z_g) + I_a Z_d + I_h Z_f = 0 \quad (2.11)$$

$$E_4 - I_h(Z_h + Z_f + Z_e) + I_g Z_f + I_b Z_e = 0 \quad (2.12)$$

Thus (2.9) represents the sum of voltages around the first loop, or mesh, of the network. The current I_a may be thought of as flowing *around* the first mesh of the circuit through the impedance $(Z_a + Z_c + Z_d)$, and if the second and third meshes were open the voltage equation for the first mesh would be given by the first two terms of (2.9), *viz.*,

$$E_1 - I_a(Z_a + Z_c + Z_d) = 0$$

Similarly, I_b may be considered to flow *around* the second mesh, etc. Thus when all these "mesh" currents are flowing, (2.9) states that the applied voltage of mesh 1 is equal to the drop around mesh 1 due to its mesh current I_a , plus the drop in Z_c (the common element between mesh 1 and mesh 2) due to the mesh current I_b of mesh 2, plus the drop in Z_d (which is common to mesh 1 and mesh 3) due to the mesh current I_g of mesh 3.

Thus, when a current is considered to flow around a mesh rather than only through one branch, the current is called a "mesh current" and is usually designated by a numerical subscript corresponding to a particular mesh. If the upper left-hand mesh in Fig. 2.1 is denoted by the number 1, the mesh current for this mesh is I_1 . For circuit of Fig. 2.1,

$$I_1 = I_a \quad (2.13) \quad I_3 = I_g \quad (2.15)$$

$$I_2 = I_b \quad (2.14) \quad I_4 = I_h \quad (2.16)$$

It is possible therefore to write the network voltage equations using either the branch currents or the mesh currents. Since there are four meshes in the network of Fig. 2.1, four mesh currents completely determine the currents through all elements of the network.

Referring to (2.13) and (2.9), and to Fig. 2.1, the total impedance in mesh 1 through which current I_1 flows is designated as Z_{11} . Thus, for Fig. 2.1,

$$Z_{11} = Z_a + Z_c + Z_d \quad (2.17)$$

Similarly,

$$Z_{22} = Z_b + Z_e + Z_c \quad (2.18)$$

$$Z_{33} = Z_d + Z_f + Z_g \quad (2.19)$$

$$Z_{44} = Z_h + Z_f + Z_e \quad (2.20)$$

Common elements between two meshes of the network are designated by a similar double-subscript notation. Thus Z_c is common to meshes 1 and 2, is the impedance in mesh 1 through which I_2 flows, and is designated as Z_{12} . Since the impedance in mesh 2 through which current I_1 flows is the same as Z_{12} , $Z_{21} = Z_{12}$. If the positive directions of the mesh currents I_1 and I_2 are in the same direction through Z_c , $Z_{12} = Z_c$. If the positive directions of mesh currents are in opposite directions through Z_c , the voltages across Z_c due to these currents are opposite in sign, and $Z_{12} = -Z_c$. Thus, for mesh currents all assumed clockwise as in Fig. 2.1,

$$Z_{12} = -Z_c \quad (2.21) \qquad Z_{34} = -Z_f \quad (2.23)$$

$$Z_{24} = -Z_e \quad (2.22) \qquad Z_{31} = -Z_d \quad (2.24)$$

Consider the circuit of Fig. 2.2.

$$Z_{11} = Z_g + Z_1 + Z_3 \quad (2.25)$$

$$Z_{22} = Z_2 + Z_L + Z_3 \quad (2.26)$$

$$Z_{12} = Z_{21} = -Z_3 \quad (2.27)$$

for the mesh current directions indicated in the figure.

The impedance presented to the generator at the terminals C and D is

$$Z_{CD} = Z_1 + \frac{Z_3(Z_2 + Z_L)}{Z_3 + Z_2 + Z_L}$$

Figure 2.2 is a special case of the more general two-mesh network of Fig. 2.3. Here the dotted rectangle represents a generalized four-terminal network connected between the generator and load impedance Z_L . In practice, the connecting system might be a

circuit of the form shown in Fig. 2.2, or it might be a transformer, a filter, a transmission line, or a system of antennas. The network equations are applicable to a wide variety of problems.

For the system shown in Fig. 2.3 some of the elements of the network may be unknown so that it is necessary to define the mesh impedances in terms of an operational procedure. In Fig. 2.2 the impedance $(Z_{CD})_{EF \text{ open}} = Z_1 + Z_3$ can be measured by discon-

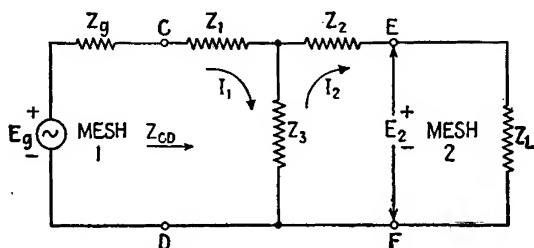


FIG. 2.2.—Example of a two-mesh network, mesh currents indicated.

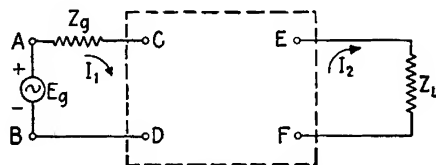


FIG. 2.3.—Generalized two-mesh network.

necting the load Z_L and measuring the impedance Z_{CD} with the terminals E and F open. By using this procedure, it is possible to define Z_{11} for Fig. 2.3 as

$$Z_{11} = Z_g + (Z_{CD})_{EF \text{ open}}$$

The symbol $(Z_{CD})_{EF \text{ open}}$ means the impedance measured at the terminals C and D with E and F open, or with $Z_L = \infty$. Similarly,

$$Z_{22} = Z_L + (Z_{EF})_{CD \text{ open}}$$

In the circuit of Fig. 2.2 it is evident that when a current flows around the first mesh a voltage will appear across the impedance Z_3 and hence in the second mesh. This voltage might be measured as the open-circuit terminal voltage across E and F . Let $(E_2)_{EF \text{ open}}$ be this open-circuit voltage (*i.e.*, the value of E_2 in Fig. 2.2 with $Z_L = \infty$), for a current I_1 in the first mesh. Then

$$(E_2)_{EF \text{ open}} = I_1 Z_3 \quad (2.28)$$

With the assumed positive directions in Fig. 2.2, $Z_{12} = -Z_3$, and $Z_{12} = -(E_2)_{EF \text{ open}}/I_1$ from (2.28). Since the currents and voltages have the same designations in Figs. 2.3 and 2.2, Z_{12} for the network of Fig. 2.3 can be defined as $Z_{12} = -(E_2)_{EF \text{ open}}/I_1$.

It is customary to write the voltage equation (2.9) in the "mesh" form

$$E_\theta - I_1 Z_{11} - I_2 Z_{12} \cdots - I_n Z_{1n} \cdots = 0 \quad (2.29)$$

as the externally applied voltage E_θ is assumed positive when acting in the direction of I_1 and all the other terms in the equation are given negative signs. Equation (2.29) is a general equation that can be used for any complicated system of meshes in which each term represents the voltage produced in the first mesh by current flow in each of the other meshes. For the circuit of Fig. 2.2, the voltage equations become

$$E_\theta - I_1 Z_{11} - I_2 Z_{12} = 0 \quad (2.30)$$

$$0 - I_1 Z_{12} - I_2 Z_{22} = 0 \quad (2.31)$$

where the impedances are given by (2.25), (2.26), and (2.27). These equations may be written in the alternate forms

$$E_\theta = I_1 Z_{11} + I_2 Z_{21} \quad (2.32)$$

$$0 = I_1 Z_{12} + I_2 Z_{22} \quad (2.33)$$

From (2.31),

$$I_2 = -I_1 \frac{Z_{12}}{Z_{22}} \quad (2.34)$$

Substituting this value for I_2 in (2.30),

$$E_\theta - I_1 Z_{11} + I_1 \frac{Z_{12} Z_{21}}{Z_{22}} = 0 \quad (2.35)$$

and since $Z_{12} = Z_{21}$, (2.35) becomes

$$E_\theta - I_1 \left(Z_{11} - \frac{Z_{12}^2}{Z_{22}} \right) = 0 \quad (2.36)$$

$$Z_{AB} = \frac{E_\theta}{I_1} = Z_{11} - \frac{Z_{12}^2}{Z_{22}} \quad (2.37)$$

Equation (2.37) is very useful since it expresses the input impedance of the two-mesh network when there are no sources of emf in the second mesh. From (2.36),

$$I_1 = \frac{E_\theta}{Z_{11} - \frac{Z_{12}^2}{Z_{22}}} \quad (2.38)$$

whereas, from (2.34),

$$I_2 = - \frac{E_g Z_{12}}{Z_{11} Z_{22} - Z_{12}^2} \quad (2.39)$$

3. Inductive Coupling.—As an example of the application of the network equations, consider the transformer shown in Fig. 3.1. Suppose that two coils of wire are arranged as shown with the turns of each in the same direction relative to the common axis, so that a part of the magnetic field of one coil threads through the other. The current I_1 produces a voltage of self-induction $-j\omega L_1 I_1$ in the upper coil, opposing the current and hence given the minus sign. Simultaneously there is a voltage $-j\omega M I_1$ induced in L_2 . The algebraic sign attached to the magnitude of the mutual inductance depends on the geometrical arrangement of the coils in space and

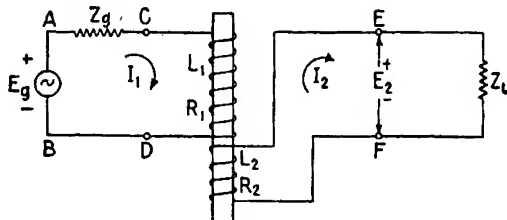


FIG. 3.1.—Two-mesh transformer circuit.

on the *assumed positive current* directions in the circuits. If the voltage induced in a “first” circuit by the current in a “second” circuit is in the same direction as the voltage of self-induction in the first circuit (the instantaneous currents in both circuits increasing in the positive direction), M is a positive number, just as L is. This means that, if the magnetic fields are *additive* when instantaneous currents are in the assumed positive direction in each coil, M is a *positive* number of henrys. When the magnetic fields are *opposed* when currents are in the assumed positive direction in each coil, M is a *negative* number of henrys.

For the coil geometry of Fig. 3.1 and the assumed positive directions of current the magnetic fields subtract and M is a negative number of henrys. The symbol is always M , although M may be algebraically positive or negative. The equation for the primary circuit is

$$E_g = I_1 Z_g + I_1 R_1 + I_1 j\omega L_1 + I_2 j\omega M \quad (3.1)$$

or

$$E_g = I_1 Z_{11} + I_2 Z_{12} \quad (3.2)$$

so that

$$Z_{12} = j\omega M \quad (3.3)$$

For the secondary circuit the equation is

$$0 = I_2 Z_L + I_2 R_2 + I_2 j\omega L_2 + I_1 j\omega M \quad (3.4)$$

or

$$I_1 Z_{21} + I_2 Z_{22} = 0 \quad (3.5)$$

and

$$Z_{21} = j\omega M = Z_{12} \quad (3.6)$$

If the assumed positive direction of I_2 were opposite to that in Fig. 3.1, the magnetic fields would add, $Z_{12} = j\omega M$ as before, but

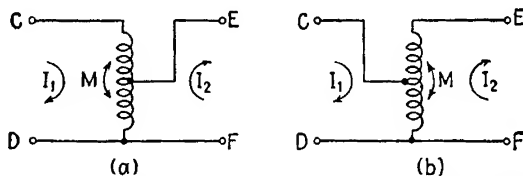


FIG. 3.2.—Tapped-coil or autotransformer connection. M is a negative number of henrys for the assumed positive directions of I_1 and I_2 .

M would be a positive number of henrys. The symbol M “carries its own sign.”

The coupling may be provided by a tapped coil or autotransformer as in Fig. 3.2. The coil might be used as the mutual element between two circuits in such a way as to make M a positive number as in Fig. 3.3, where the magnetic fields due to assumed positive currents are additive. In all these cases, $Z_{12} = Z_{21} = j\omega M$, where M is positive for additive magnetic fields and negative for opposed magnetic fields corresponding to the assumed current directions. For the circuit of Fig. 3.1, (2.37) becomes

$$Z_{AB} = Z_{11} - \frac{Z_{12}^2}{Z_{22}} = Z_0 + R_1 + j\omega L_1 + \frac{\omega^2 M^2}{Z_L + R_2 + j\omega L_2} \quad (3.7)$$

Similarly,

$$Z_{CD} = R_1 + j\omega L_1 + \frac{\omega^2 M^2}{Z_L + R_2 + j\omega L_2} \quad (3.8)$$

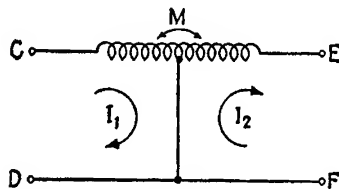


FIG. 3.3.—Tapped-coil arrangement in which M is a positive number of henrys for the assumed positive directions of I_1 and I_2 .

The input impedance to a transformer is given by (3.8). If the secondary circuit is open, $Z_L = \infty$, and

$$(Z_{CD})_{EF \text{ open}} = R_1 + j\omega L_1 \quad (3.9)$$

If the terminals E and F are short-circuited, $Z_L = 0$, and

$$(Z_{CD})_{EF \text{ shorted}} = R_1 + j\omega L_1 + \frac{\omega^2 M^2}{R_2 + j\omega L_2} \quad (3.10)$$

These relations will be used in Sec. 4.

4. Equivalent Networks.—It is often desirable to transform a four-terminal network from a T configuration, Fig. 4.1a, to an equivalent network having a Π configuration, Fig. 4.1b, and vice versa.¹ Two such networks are equivalent when they have identical terminal currents and voltages between terminals under

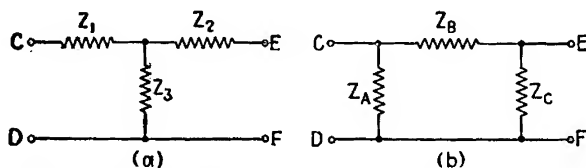


FIG. 4.1.—T and Π networks equivalent to each other when equations (4.6) to (4.12) hold.

identical conditions external to the terminals. The corresponding input impedances must be equal. With E and F left open, the impedance at the terminals CD of the Π network of Fig. 4.1b is

$$Z_{CD} = \frac{Z_A(Z_B + Z_C)}{Z_A + Z_B + Z_C} \quad (4.1)$$

and for the T network of Fig. 4.1a is

$$Z_{CD} = Z_1 + Z_3 \quad (4.2)$$

Equating (4.1) and (4.2),

$$Z_1 + Z_3 = \frac{Z_A(Z_B + Z_C)}{Z_A + Z_B + Z_C} \quad (4.3)$$

Similarly,

$$Z_2 + Z_3 = \frac{Z_C(Z_A + Z_B)}{Z_A + Z_B + Z_C} \quad (4.4)$$

Equating the impedances Z_{CD} with E and F short-circuited,

$$Z_1 + \frac{Z_2 Z_3}{Z_2 + Z_3} = \frac{Z_A Z_B}{Z_A + Z_B} \quad (4.5)$$

¹ The T network is also known as the Y, or star, connection; the Π network is also known as the Δ (delta) connection.

Eliminating Z_1 between (4.3) and (4.5) and substituting the value of $Z_2 + Z_3$ in (4.4),

$$Z_3 = \frac{Z_A Z_C}{Z_A + Z_B + Z_C} \quad (4.6)$$

whence, from (4.3),

$$Z_1 = \frac{Z_A Z_B}{Z_A + Z_B + Z_C} \quad (4.7)$$

and from (4.4),

$$Z_2 = \frac{Z_B Z_C}{Z_A + Z_B + Z_C} \quad (4.8)$$

Multiplying (4.4) and (4.5),

$$Z_1 Z_2 + Z_2 Z_3 + Z_3 Z_1 = \frac{Z_A Z_B Z_C}{Z_A + Z_B + Z_C} \quad (4.9)$$

Dividing (4.9) by (4.8), (4.6), (4.7) gives (4.10), (4.11), (4.12), for the values of the impedances of a Π network equivalent to a given T.

$$Z_A = \frac{Z_1 Z_2 + Z_2 Z_3 + Z_3 Z_1}{Z_2} \quad (4.10)$$

$$Z_B = \frac{Z_1 Z_2 + Z_2 Z_3 + Z_3 Z_1}{Z_3} \quad (4.11)$$

$$Z_C = \frac{Z_1 Z_2 + Z_2 Z_3 + Z_3 Z_1}{Z_1} \quad (4.12)$$

Since impedances of circuit elements are functions of frequency, two networks constructed to be equivalent at one frequency may

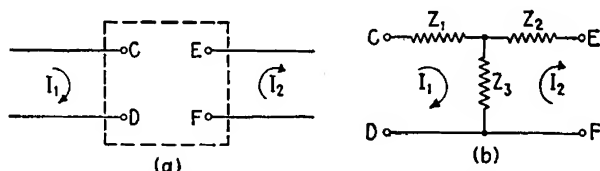


FIG. 4.2.—(a) General four-terminal network; (b) its equivalent T when (4.22) to (4.24) hold, as described in the text.

not be equivalent at any other frequency. It is not always possible to construct physically an equivalent network using ordinary circuit elements since some resistance values may come out negative.

It is useful to develop the expressions for the T or Π network equivalent to a general four-terminal network in terms of the input impedances of the general network under open- and short-circuit conditions. Figure 4.2a shows a four-terminal network. Let

$$(Z_{CD})_{EF \text{ open}} = Z_{O1} \quad (4.13) \quad (Z_{EF})_{CD \text{ open}} = Z_{O2} \quad (4.14)$$

and

$$(Z_{CD})_{EF \text{ shorted}} = Z_{S1} \quad (4.15) \quad (Z_{EF})_{CD \text{ shorted}} = Z_{S2} \quad (4.16)$$

Then for the T of Fig. 4.2b to be equivalent it is necessary that

$$Z_1 + Z_3 = Z_{O1} \quad (4.17)$$

and

$$Z_2 + Z_3 = Z_{O2} \quad (4.18)$$

The short-circuited impedances must also be equal,

$$Z_{S1} = Z_1 + \frac{Z_2 Z_3}{Z_2 + Z_3} = \frac{Z_1 Z_2 + Z_2 Z_3 + Z_3 Z_1}{Z_2 + Z_3} \quad (4.19)$$

and

$$Z_{S2} = Z_2 + \frac{Z_1 Z_3}{Z_1 + Z_3} = \frac{Z_1 Z_2 + Z_2 Z_3 + Z_3 Z_1}{Z_1 + Z_3} \quad (4.20)$$

Substituting (4.18) in (4.19) and (4.17) in (4.20),

$$Z_1 Z_2 + Z_2 Z_3 + Z_3 Z_1 = Z_{O2} Z_{S1} = Z_{O1} Z_{S2} \quad (4.21)$$

Substituting $Z_1 = Z_{O1} - Z_3$ from (4.17) and $Z_2 = Z_{O2} - Z_3$ from (4.18) in (4.21),

$$Z_3 = \pm \sqrt{Z_{O2}(Z_{O1} - Z_{S1})} = \pm \sqrt{Z_{O1}(Z_{O2} - Z_{S2})} \quad (4.22)$$

whence

$$Z_1 = Z_{O1} \mp \sqrt{Z_{O2}(Z_{O1} - Z_{S1})} \quad (4.23)$$

and

$$Z_2 = Z_{O2} \mp \sqrt{Z_{O2}(Z_{O1} - Z_{S1})} \quad (4.24)$$

The choice of the + or - sign in (4.22) to (4.24) is determined by the phase of the voltage at the terminals EF with respect to the current at the terminal C and must be chosen to satisfy (2.28). Thus the determination of Z_3 requires the measurement or calculation of the phase relationship between $(E_2)_{EF \text{ open}}$ and I_1 . Where phase relationships may be ignored and only the input and output impedances must be equivalent, either choice of sign suffices. Of course, if the + sign is used in (4.22), the minus sign must be used in (4.23) and (4.24), and vice versa.

As an example of the application of these equations, let it be required to find the T network equivalent to the transformer of Fig. 3.1. From (3.9) and (3.10),

$$Z_{O1} = R_1 + j\omega L_1$$

$$Z_{s1} = R_1 + j\omega L_1 + \frac{\omega^2 M^2}{R_2 + j\omega L_2}$$

Similarly,

$$Z_{o2} = R_2 + j\omega L_2$$

Substituting these values in (4.22), (4.23), (4.24), the minus sign being chosen in accordance with the required phase relationship of (2.28),

$$Z_3 = -j\omega M = j\omega |M| \quad (4.25)$$

$$Z_1 = R_1 + j\omega(L_1 + M) = R_1 + j\omega(L_1 - |M|) \quad (4.26)$$

$$Z_2 = R_2 + j\omega(L_2 + M) = R_2 + j\omega(L_2 - |M|) \quad (4.27)$$

It should be recalled at this point that M is a negative number of henrys for the circuit of Fig. 3.1 owing to the directions of the windings and the assumed positive directions of current. The equivalent T has the form shown in Fig. 4.3. The network of Fig. 4.3 is physically realizable with coils so long as L_1 and L_2 are each larger than $|M|$. This will always be true for a one-to-one transformer. If the transformer is step-up or step-down, one of the series reactances in the equivalent T may be negative. In constructing the equivalent circuit, the negative reactance may be furnished by a series capacitor in the required branch and by making $-1/\omega C = \omega(L - |M|)$ when $|M|$ is greater than L . After C is thus chosen, the circuit is equivalent to the actual transformer at only one frequency. If the transformer is a one-to-one transformer, $L_1 = L_2 = L$. Introducing the coefficient of coupling k where

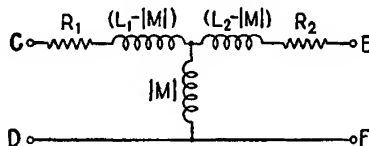


FIG. 4.3.—Equivalent T for the transformer of Fig. 3.1.

$$k = \left| \frac{M}{\sqrt{L_1 L_2}} \right| \quad (4.28)$$

then

$$Z_1 = R_1 + j\omega L(1 - k) \quad (4.29)$$

$$Z_2 = R_2 + j\omega L(1 - k) \quad (4.30)$$

$$Z_3 = j\omega Lk \quad (4.31)$$

Further development of the equivalent circuits for a transformer is given in Chap. XIII.

5. Network Theorems.—There are a number of general principles or theorems that may be employed to determine the currents:

and voltages in a network. Some of these principles are presented here without proof, which may be found in the literature.

Superposition Theorem.—Suppose that an active network has several generators introducing voltages at various points in the network and that it is required to find the current through a particular element when all these generators are operating. The superposition theorem states that the instantaneous current through any circuit element is the sum of the instantaneous currents caused by each generator acting separately. It is necessary that all the circuits be unchanged when the current due to a given generator is being calculated; therefore other generators must be considered to be “shut down” but not removed from the circuit, so that the internal impedances of all generators remain in the circuit at all times. This last condition means that each generator except the one for which the current is being calculated is replaced by its internal impedance.

The superposition theorem holds because of the linearity of the network elements, since in a linear system of differential equations the sum of any number of solutions is also a solution.

Reciprocity Theorem.—Suppose that a voltage is applied in series with a certain branch of a network and the resulting current in any branch is measured. The ratio of voltage introduced to current flowing will have a certain value. The reciprocity theorem states that, if the positions of generator and ammeter are interchanged, the ratio of voltage to current is the same. Again, for this theorem to apply, the impedances of the several branches must not be altered when the generating and measuring devices are interchanged. The reciprocal properties of a network are an indication of the bilateral nature of linear networks.

Thévenin's Theorem.—Consider an active network of generators and impedances in any configuration whatever, having two terminals. Thévenin's theorem states that in the steady state the system may be replaced by a simple circuit containing a source of voltage in series with an impedance. The voltage of the source is the open-circuit voltage across the two terminals of the original network, and the series impedance is the impedance “looking in” at the terminals with all the generators of the original system replaced by their internal impedances. As an example, consider the circuit of Fig. 5.1a, which is a radio-frequency amplifier with a tuned plate load. The problem is to reduce the whole system to a simple series circuit so far as small variational, or alternating,

currents are concerned, so that its frequency-response characteristics may be studied more easily. By the use of the equivalent plate-circuit theorem, developed in Chap. XI, the circuit may be reduced to the form of Fig. 5.1b. By applying Thévenin's theorem to that part of the circuit to the left of terminals AB , the whole circuit can be reduced to Fig. 5.1c, where E' is the open-circuit

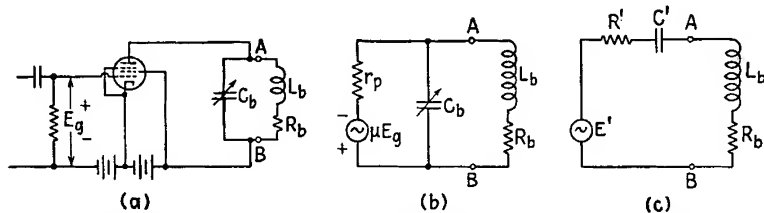


FIG. 5.1.—Reduction of a vacuum-tube amplifier circuit to a simple equivalent series circuit by means of Thévenin's theorem.

voltage across AB and $(R' - j1/\omega C')$ is the impedance looking to the left at AB . With R_b and L_b disconnected in Fig. 5.1b, the open-circuit voltage at AB is

$$(E_{AB})_{\text{open}} = -\mu E_g \frac{-j \frac{1}{\omega C_b}}{r_p - j \frac{1}{\omega C_b}} = -\frac{\mu E_g}{1 + j r_p \omega C_b} \quad (5.1)$$

The impedance looking to the left at AB is the impedance of C_b and r_p in parallel. Therefore,

$$Z_{AB} = \frac{r_p \left(-j \frac{1}{\omega C_b} \right)}{r_p - j \frac{1}{\omega C_b}} = \frac{r_p}{1 + j r_p \omega C_b}$$

Rationalizing Z_{AB} ,

$$Z_{AB} = \frac{r_p - j r_p^2 \omega C_b}{1 + r_p^2 \omega^2 C_b^2} = R' - j \frac{1}{\omega C'} \quad (5.2)$$

where

$$R' = \frac{r_p}{1 + r_p^2 \omega^2 C_b^2} \quad (5.3)$$

and

$$C' = C_b \frac{1 + r_p^2 \omega^2 C_b^2}{r_p^2 \omega^2 C_b^2} \quad (5.4)$$

In many applications,

$$r_p^2 \omega^2 C_b^2 \gg 1$$

so that

$$R' \doteq \frac{1}{r_p \omega^2 C_b^2} \quad (5.5)$$

and

$$C' \doteq C_b \quad (5.6)$$

The effect of the tube on the tuned circuit is easily calculated from (5.3) and (5.4). The presence of the tube increases the effective series resistance of the tuned circuit, lowering the Q and broadening the response curve.

When the impedances of part of a network are independent of frequency, the calculation of transients sometimes may be simplified by applying Thévenin's theorem to that part.

Conditions for the Transfer of Maximum Power.—At any frequency, a network containing generators and impedances can be replaced (by Thévenin's theorem) by a simple series circuit containing a generator and a series impedance. Likewise, a network not containing generators may be replaced by its input impedance. If a generator and a load impedance are connected together and the generator impedance is fixed, the condition for maximum power transfer is that the load impedance be the conjugate of the internal impedance of the generator.

Figure 5.2a shows the generalized circuit and its equivalent, Fig. 5.2b, where E_1 is equal to $(E_{AB})_{\text{open}}$, Z_1 is the internal impedance of network 1, and Z_2 is the input impedance of network 2. Referring to Fig. 5.2b, the power transferred across terminals CD into R_2 is

$$P_{CD} = |E_{CD}| |I| \cos \theta \quad (5.7)$$

where

$$|E_{CD}| = |I| \sqrt{R_2^2 + X_2^2} \quad (5.8)$$

and

$$\cos \theta = \frac{R_2}{\sqrt{R_2^2 + X_2^2}} \quad (5.9)$$

so that

$$P_{CD} = |I|^2 R_2 \quad (5.10)$$

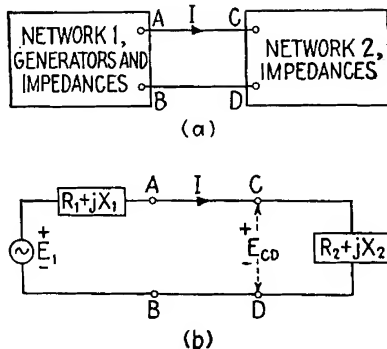


FIG. 5.2.—(a) General form of two coupled networks; (b) its equivalent form for the calculation of power transfer.

But the current $|I|$ is

$$|I| = \frac{|E_1|}{|Z_1 + Z_2|} = \frac{|E_1|}{\sqrt{(R_1 + R_2)^2 + (X_1 + X_2)^2}} \quad (5.11)$$

and the power transferred is

$$P_{cd} = \frac{|E_1|^2 R_2}{(R_1 + R_2)^2 + (X_1 + X_2)^2} \quad (5.12)$$

Suppose now that the generating-system impedance is fixed and the problem is to find the load impedance which will absorb the maximum power from the generator. So far as the reactance X_2 is concerned the largest value of current in (5.11) is secured by making the term $(X_1 + X_2)$ equal to zero. Therefore,

$$X_2 = -X_1 \quad (5.13)$$

is the reactance condition for maximum power transfer. If this condition is met, (5.12) reduces to

$$P_{cd} = \frac{|E_1|^2 R_2}{(R_1 + R_2)^2} \quad (5.14)$$

Differentiating (5.11) with respect to R_2 and setting the result equal to zero,

$$R_2 = R_1 \quad (5.15)$$

This is the resistance condition for maximum power transfer.

6. Impedance Transformation.—In radio and communication circuits the generators are vacuum tubes, microphones, photocells, or other sources of voltage. The design of these generators is controlled usually by considerations other than internal impedance. A vacuum tube often has a high internal impedance. On the other hand, the common loads to which power is to be transferred are antennas, or lines, or loudspeakers, or other loads, many of which are inherently of low input impedance. The problem arises of transforming impedances by means of networks. Such impedance-transforming networks often are called matching sections. The general problem is to take a given load impedance and design a network which transforms that load impedance to any desired value at the input terminals of the matching device.

In many radio-frequency applications of matching devices the band width is relatively narrow and tuned circuits ordinarily are used in the matching networks. On the other hand, in audio- or

video-frequency circuits the relative frequency band width is very large so that other methods must be employed. Matching may be done to achieve maximum power transfer into the load or to provide any desired load for a given device. In many cases the desired load is not the load that would absorb the maximum power.

In general, it is desirable to keep the losses in the matching section to a minimum so that as much as possible of the power transferred to the matching section is handed on to the load.

7. Radio-frequency Transforming and Matching Networks.—

A very important example of an impedance-transforming network is the parallel-resonant circuit known as a “tank” circuit. Figure 7.1 shows a typical tank circuit used as the plate load for a vacuum

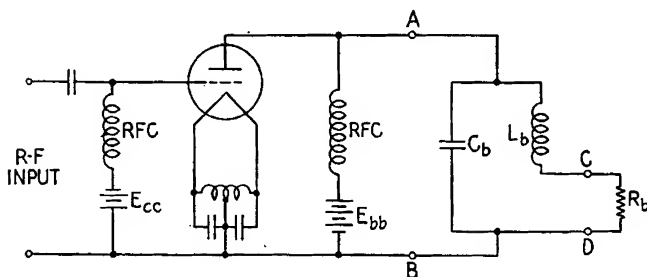


FIG. 7.1.—A power tube with a tuned load. The tank circuit acts like an impedance-transforming circuit with terminals AB and CD .

tube designed to deliver considerable power. C_b and L_b are the tuning capacitance and inductance and R_b is the load resistance in which the radio-frequency power output of the tube will be dissipated. The resistance of the load to which radio-frequency power is to be supplied and the resistance of the coil L_b are included in R_b , so that

$$R_b = R_{\text{coil}} + R_{\text{load}} \quad (7.1)$$

The tank circuit achieves two important results. First, it serves as a transforming network to transform a relatively low impedance load such as a line or an antenna to a higher impedance that is purely resistive and that is required for good performance of the tube. Second, it acts as a filter circuit to prevent the higher harmonic components of the plate current from developing appreciable voltages in the load circuit. This latter function is particularly important where the vacuum tube is operating as a Class C amplifier, whose plate current is rich in harmonic content.

The admittance of the parallel $L_b C_b$ circuit is

$$Y_{AB} = Y_L = j\omega C_b + \frac{R_b}{R_b^2 + \omega^2 L_b^2} - j \frac{\omega L_b}{R_b^2 + \omega^2 L_b^2} \quad (7.2)$$

In order that the impedance presented to the tube by the tank circuit shall be pure resistance at the fundamental frequency, the j terms in (7.2) must vanish, or

$$j\omega C_b - j \frac{\omega L_b}{R_b^2 + \omega^2 L_b^2} = 0 \quad (7.3)$$

This condition leads to the following expressions for parallel resonance:

$$R_b^2 + \omega^2 L_b^2 = \frac{L_b}{C_b} \quad (7.4)$$

$$\omega^2 = \frac{1}{L_b C_b} - \frac{R_b^2}{L_b^2} \quad (7.5)$$

If (7.4) is satisfied, the admittance becomes

$$Y_L = \frac{R_b C_b}{L_b} \quad (7.6)$$

and the impedance is given by the reciprocal of (7.6), or

$$Z_L = R_L = \frac{L_b}{C_b R_b} \quad (7.7)$$

It is possible, therefore, to design the tank circuit to have any desired impedance R_L within certain limits. Suppose that the frequency ω is given, that the load impedance R_b is fixed, and that the desired plate-load resistance R_L has been chosen. The value of L_b/C_b thus is fixed by the given conditions, as is the frequency, so that L_b and C_b are not independent and one of them may be eliminated from the equations.

Substituting (7.4) in (7.7),

$$R_L = \frac{L_b}{C_b R_b} = \frac{R_b^2 + \omega^2 L_b^2}{R_b} \quad (7.8)$$

Dividing out R_b^2 from the numerator,

$$R_L = R_b \left(1 + \frac{\omega^2 L_b^2}{R_b^2} \right) = R_b (1 + Q_b^2) \quad (7.9)$$

where Q_b is the quality factor of the inductive branch, including the load resistance, at the parallel-resonance frequency, which is also

the impressed frequency. The ratio of tank impedance R_L to load impedance R_b is therefore

$$\frac{R_L}{R_b} = 1 + Q_b^2 \quad (7.10)$$

If Q_b is large compared with unity, the ratio is approximately equal to Q_b^2 . It is evident that the range of possible impedance ratios, achievable with this circuit, is limited by the possible range of values of Q_b . Since R_b includes both load resistance and coil resistance, the highest attainable value of Q_b is always less than the Q of the coil alone.

If the tank circuit is to be used for transfer of radio-frequency power from the plate circuit to the load, it should have a high efficiency. The efficiency of the tank circuit as a transducer between plate circuit and load is given by

$$\text{Efficiency} = \frac{\text{output}}{\text{input}} = \frac{|I|_{\text{coil}}^2 R_{\text{load}}}{|I|_{\text{coil}}^2 R_b} = \frac{R_{\text{load}}}{R_b} \quad (7.11)$$

and, from (7.1),

$$\text{Efficiency} = \frac{1}{1 + \frac{R_{\text{coil}}}{R_{\text{load}}}} \quad (7.12)$$

Since

$$Q_b = \frac{\omega L}{R_b} = \frac{\omega L}{R_{\text{coil}} + R_{\text{load}}} \quad (7.13)$$

the efficiency expressed in terms of Q is

$$\text{Efficiency} = \frac{1}{\frac{Q_b}{Q_{\text{coil}}} + 1} = 1 - \frac{Q_b}{Q_{\text{coil}}} \quad (7.14)$$

For high efficiency the Q of the coil should be much greater than the Q_b of the inductive branch including the load resistance. In order to get a large ratio of impedance transformation, say 100 to 1, it is necessary to have the value of $Q_b \doteq 10$ from (7.10); and in order to get high efficiency of power transfer, say 95 per cent, it is necessary to have the Q of the coil equal to 200. Hence the upper limit to the ratio of impedance transformation is set by the available Q of the coil and the desired efficiency.

If the tank circuit is to function as a band-pass filter as well as an impedance-transforming device, the value of Q_b must not be

too low. For example, in the tank circuit of a Class *C* amplifier it is usually desirable to reduce the second harmonic voltage across the output load to a small value in comparison with the fundamental voltage. The admittance of the tank circuit for the second harmonic 2ω is

$$(Y_L)_{2nd} = 2j\omega C_b + \frac{R_b}{R_b^2 + 4\omega^2 L_b^2} - 2j \frac{\omega L_b}{R_b^2 + 4\omega^2 L_b^2} \quad (7.15)$$

Upon neglecting the real term in comparison with the imaginary terms and dropping R_b^2 , which is much less than $4\omega^2 L_b^2$ if Q is of the order of 10, (7.15) becomes

$$(Y_L)_{2nd} = 2j \left(\omega C_b - \frac{1}{4\omega L_b} \right) = j \frac{4\omega^2 L_b C_b - 1}{2\omega L_b} \quad (7.16)$$

From (7.16) and (7.5),

$$\begin{aligned} (Z_L)_{2nd} &= j \frac{2\omega L_b}{1 - 4\omega^2 L_b C_b} = j \frac{2\omega L_b}{1 - 4 + 4 \frac{R_b^2 C_b}{L_b}} \\ &= -j \frac{2\omega L_b}{3 - 4 \frac{R_b^2 C_b}{L_b}} \end{aligned} \quad (7.17)$$

and since

$$\frac{R_b^2 C_b}{L_b} \doteq \frac{1}{Q_b^2} \quad (7.18)$$

$$(Z_L)_{2nd} \doteq -j \frac{2\omega L_b}{3} \quad (7.19)$$

whereas, at the fundamental frequency, (7.7) yields

$$Z_L = \frac{L_b}{R_b C_b} \doteq \frac{\omega^2 L_b^2}{R_b} = Q_b \omega L_b \quad (7.20)$$

Then

$$|(Z_L)_{2nd}| \doteq \frac{2}{3Q_b} |Z_L| \quad (7.21)$$

In actual operation of the type of tube used in the circuit of Fig. 7.1, the second harmonic component of the plate current is less than the fundamental component. Then from (7.21) the second harmonic voltage across the load is less than $2/(3Q_b)$ times the fundamental voltage. Also, since the inductive reactance of the coil is doubled at the second harmonic frequency, the second harmonic current through the inductive branch is less than $1/(3Q_b)$ times the fundamental current. For example, if $Q_b = 11$, the second har-

monic current in the load, or inductive, branch is less than 3 per cent of the fundamental component.

To sum up the limitations for a simple tank circuit, if harmonic filtering is required, the tank circuit Q , should be kept as high as 10. Consequently, a high transformation ratio between the plate circuit and load requires that the Q of the coil alone be high for good efficiency.

It may be desirable to provide a means for changing the impedance presented to the tube. A form of tank circuit in which this is

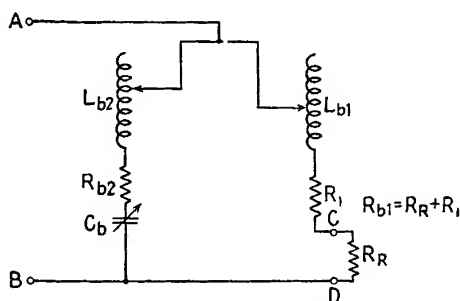


FIG. 7.2.—Adjustable tank circuit.

possible is shown in Fig. 7.2. The admittance of the plate load is given by

$$Y_{AB} = \frac{R_{b1}}{R_{b1}^2 + X_1^2} + \frac{R_{b2}}{R_{b2}^2 + X_2^2} + j \left(\frac{X_1}{R_{b1}^2 + X_1^2} + \frac{X_2}{R_{b2}^2 + X_2^2} \right) \quad (7.22)$$

where $X_1 = \omega L_1$ and $X_2 = \omega L_2 - 1/\omega C_b$. The condition that the plate load be pure resistance requires that the j terms in (7.22) vanish, or that

$$\frac{X_1}{R_{b1}^2 + X_1^2} = - \frac{X_2}{R_{b2}^2 + X_2^2} \quad (7.23)$$

whence

$$\frac{X_1}{X_2} = - \frac{R_{b1}^2 + X_1^2}{R_{b2}^2 + X_2^2} = - \frac{|Z_1|^2}{|Z_2|^2} \quad (7.24)$$

If the values of L_1 and L_2 are adjusted for parallel resonance of the circuit, which is the same condition as making the plate load a pure resistance,

$$Y_{AB} = G_L = \frac{R_{b1}}{R_{b1}^2 + X_1^2} + \frac{R_{b2}}{R_{b2}^2 + X_2^2} \quad (7.25)$$

and since $R_{b2}^2 + X_2^2 = (-X_2/X_1)(R_{b1}^2 + X_1^2)$ from (7.24),

$$G_L = \frac{R_{b1} - \frac{X_1}{X_2} R_{b2}}{R_{b1}^2 + X_1^2} \quad (7.26)$$

and

$$R_L = \frac{R_{b1}^2 + X_1^2}{R_{b1} - \frac{X_1}{X_2} R_{b2}} \quad (7.27)$$

If the Q of each branch is high, $R_{b1}^2 \ll X_1^2$, $-X_1/X_2 \doteq 1$, (7.27) becomes

$$R_L \doteq \frac{\omega^2 L_{b1}^2}{R_{b1} + R_{b2}} \quad (7.28)$$

Thus by varying the taps on the two coils it is possible to adjust ωL_{b1} to vary R_L over a range of values and then tune the circuit by varying ωL_{b2} . This system of "matching," or transforming, a given load into an impedance suitable for the tube is rather poor for harmonic suppression since there is no capacitance path to by-pass the higher harmonic components of plate current to ground. More elaborate matching devices between vacuum-tube generators and low impedance loads are discussed in Chap. XIV.

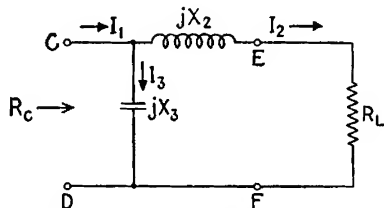


FIG. 7.3.—Matching section for which the input resistance R_c is greater than the load resistance R_L when properly adjusted.

From (7.10) it is evident that low ratios of transformation mean low ratios of reactance to resistance. This in turn means that special arrangements must be used for harmonic filtering. However, since transmission lines and half-wave antennas are low impedance devices, there is need for matching sections whose ratio of impedance transformation is low. Figure 7.3 shows a typical matching section of pure reactance elements for transforming the load resistance R_L to a larger value R_c . It is evident from the figure that this is the same circuit as the tank circuit already considered, Fig. 71. Applying (7.7) to the circuit of Fig. 7.3,

$$R_c = \frac{L}{CR_L} = \frac{\omega L}{\omega CR_L} = \frac{-X_2 X_3}{R_L} \quad (7.29)$$

Suppose, however, that it is assumed that X_2 and X_3 of Fig. 7.3 are reactances of unknown sign and that R_L and R_c , the desired impedance at CD , are known. Then, setting R_c equal to the input impedance at CD ,

$$R_c = \frac{jX_3(R_L + jX_2)}{R_L + jX_2 + jX_3} \quad (7.30)$$

whence

$$R_c R_L + jX_2 R_c + jX_3 R_c = jX_3 R_L - X_2 X_3 \quad (7.31)$$

Equating the real parts of (7.31),

$$R_c R_L = -X_2 X_3 \quad (7.32)$$

which is the same equation as (7.29). However, since R_c and R_L are both positive, X_2 and X_3 must be of opposite sign. Equating the imaginary parts of (7.31),

$$X_3(R_c - R_L) = -X_2 R_c \quad (7.33)$$

Multiplying (7.33) by (7.32),

$$X_2 = \pm \sqrt{R_L(R_c - R_L)} \quad (7.34)$$

and by substitution in (7.32)

$$X_3 = \mp \frac{R_L R_c}{\sqrt{R_L(R_c - R_L)}} \quad (7.35)$$

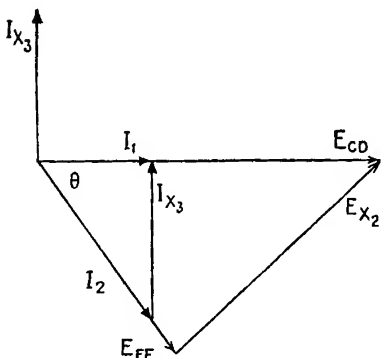


FIG. 7.4.—Phase vector diagram for the circuit of Fig. 7.3.

Thus the values of X_2 and X_3 are expressed in terms of the desired

resistance terminations. Furthermore, the \pm signs in (7.34) and (7.35) indicate that X_2 could be either inductive or capacitive. Having chosen the sign of X_2 , however, X_3 must be of the opposite sign. The particular arrangement of Fig. 7.3 always will make R_c larger than R_L . Both (7.34) and (7.35) become imaginary when $R_c < R_L$. Figure 7.4 shows the voltage and current phase-vector diagram for Fig. 7.3. The voltage E_{CD} is always greater than E_{EF} , while the current I_1 is always less than I_2 ; the product $I_1 E_{CD}$ is equal to $I_2 E_{EF}$, as would be expected in an ideal transformer. However, the input and output voltages differ in phase by an angle that approaches 90° as the ratio of R_c to R_L is increased. For the circuit of Fig. 7.3, where inductance is in series with R_L , the output voltage lags the input voltage by an angle given by

$$\cos \theta = \sqrt{\frac{R_L}{R_c}}$$

assuming no losses in L or C .

If the input resistance R_c is to be less than R_L , the matching network is turned end for end as in Fig. 7.5. Similar analysis leads to the values of X_1 and X_3 given by

$$X_1 = \pm \sqrt{R_c(R_L - R_c)} \quad (7.36)$$

$$X_3 = \pm \frac{R_L R_c}{\sqrt{R_c(R_L - R_c)}} \quad (7.37)$$

As before,

$$R_c R_L = -X_1 X_3 = \frac{L}{C} \quad (7.38)$$

Again the signs in (7.36) and (7.37) could be chosen to place the inductance in either position as required by conditions. Figure 7.6

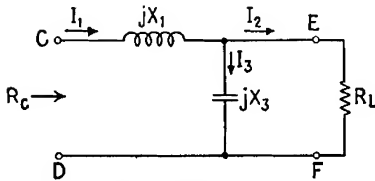


FIG. 7.5.—Matching section for which $R_c < R_L$.

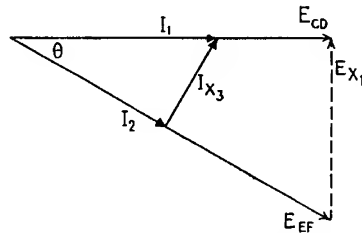


FIG. 7.6.—Phase vector diagram for the circuit of Fig. 7.5.

shows the phase vector diagram for the network of Fig. 7.5. For this case the phase lag of the output voltage with respect to the input voltage is given by

$$\cos \theta = \sqrt{\frac{R_c}{R_L}} \quad (7.39)$$

If the load impedance is not a pure resistance but contains some reactance, it is necessary merely to add a reactance in series with the load equal to the negative of the load reactance. In the circuit of Fig. 7.3 this reactance may be combined with X_2 from (7.34) and one coil or capacitor used to obtain the total required reactance.

When the wavelengths are sufficiently short, matching sections can be constructed, using sections of transmission line. These line sections have low losses and may be adjusted easily and accurately.

Their use and design are discussed in Chap. I of the reference given below.¹

If there is no dissipation in the matching section, all power transferred across the terminals CD of Figs. 7.3 and 7.5 must also be transferred to the load. It is then true that if R_c is equal to the internal impedance of a constant-voltage generator connected to the terminals CD of Figs. 7.3 and 7.5, there is a maximum power transfer across CD and across EF , and the impedance looking back into the matching section at EF (toward the generator) is equal to R_L . If, on the other hand, there is dissipation in the elements of the matching section, it is necessary in general to arrange to match at each end of the section.

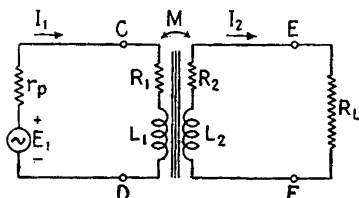


FIG. 7.7.—Schematic diagram of an iron-core transformer.

All the matching networks thus far considered were designed for a single frequency or a relatively narrow band of frequencies. In public-address systems, where it is necessary to match from vacuum-tube generators to lines or loudspeakers, and in telephone systems,

where the match is from lines to amplifiers and back to lines, the frequency bands may comprise several octaves. In a high-fidelity public-address system the frequency range is from 30 to 10,000 cps or more, perhaps nine octaves. The matching systems used are iron-core transformers having very close coupling between primary and secondary coils. Figure 7.7 shows a transformer used to transfer audio-frequency power from a vacuum-tube generator to a low-impedance load assumed to be a pure resistance R_L . From (3.8), the impedance presented to the vacuum tube by the loaded transformer when $Z_L = R_L$ is

$$Z_{CD} = R_1 + j\omega L_1 + \frac{\omega^2 M^2}{(R_2 + R_L) + j\omega L_2} \quad (7.40)$$

Rationalizing the last term and combining real and j terms,

$$Z_{CD} = R_1 + \frac{(R_2 + R_L)\omega^2 M^2}{(R_2 + R_L)^2 + \omega^2 L_2^2} + j\omega \left(L_1 - \frac{L_2 \omega^2 M^2}{(R_2 + R_L)^2 + \omega^2 L_2^2} \right) \quad (7.41)$$

¹ KING, R. W. P., H. R. MIMNO, and A. H. WING, "Transmission Lines, Antennas, and Wave Guides," McGraw-Hill Book Company, Inc., 1945.

If M is set equal to $k\sqrt{L_1 L_2}$, where k is the coefficient of coupling between coils, and if $(R_2 + R_L)$ is called R_L' , (7.41) becomes

$$Z_{CD} = R_1 + \frac{R_L' \omega^2 k^2 L_1 L_2}{R_L'^2 + \omega^2 L_2^2} + j\omega L_1 \left(1 - \frac{k^2 \omega^2 L_2^2}{R_L'^2 + \omega^2 L_2^2} \right) \quad (7.42)$$

The ideal matching section between tube and resistance load should be independent of frequency and should present a pure resistance to the generator. If ωL_2 is much larger than R_L' , this requirement is partly met. With

$$\omega L_2 \gg R_L' \quad (7.43)$$

(7.42) becomes

$$Z_{CD} = R_1 + k^2 \frac{L_1}{L_2} R_L' + j\omega L_1 (1 - k^2) \quad (7.44)$$

and if the coefficient of coupling approaches the value 1,

$$1 - k^2 \ll 1 \quad (7.45)$$

and

$$Z_{CD} \doteq R_{CD} \doteq R_1 + \frac{L_1}{L_2} R_L' \quad (7.46)$$

Under these conditions, (7.46) is independent of frequency.

Since the transformer is handling audio power, it is desirable to make its losses as small as possible. When

$$R_1 \ll \frac{L_1}{L_2} R_L' \quad (7.47)$$

and

$$R_2 \ll R_L \quad (7.48)$$

R_{CD} becomes

$$R_{CD} = \frac{L_1}{L_2} R_L = \frac{1}{\alpha^2} R_L \quad (7.49)$$

where α is given by

$$\alpha = \sqrt{\frac{L_2}{L_1}} \quad (7.50)$$

Equation (7.49) defines a condition secured by an "ideal" transformer. Since the inductance of a coil of very closely coupled turns is proportional to the square of the number of turns, (7.49) may be written

$$R_{CD} = \frac{L_1}{L_2} R_L \doteq \frac{N_1^2}{N_2^2} R_L \quad (7.51)$$

This is a very useful approximation. If the losses in the transformer are negligible,

$$I_1^2 R_{CD} = I_2^2 R_L \quad (7.52)$$

and

$$\frac{E_{CD}^2}{R_{CD}} = \frac{E_{EF}^2}{R_L} \quad (7.53)$$

whence, from (7.50) and (7.52),

$$\frac{E_{CD}}{E_{EF}} = \frac{N_1}{N_2} = \frac{I_2}{I_1} = \frac{1}{\alpha} \quad (7.54)$$

Equation (7.54) shows the relationship between primary and secondary voltages and currents for an ideal transformer.

The conditions which must be fulfilled to make an actual transformer approach the ideal are, first, that ωL_2 be much larger than R_L' , (7.43); second, that $k \doteq 1$, (7.45); third, that R_1 and R_2 be negligible as stated in (7.47) and (7.48); fourth, that effects of stray capacitance between turns and coils be negligible.

The first condition cannot be met at low frequencies since ωL_2 is small. The second condition cannot be met at high frequencies since the term $j\omega L_1(1 - k^2)$ becomes appreciable as ω increases even when $1 - k^2$ is small. The fourth condition cannot be met at very high frequencies since the effects of stray capacitance increase with frequency. As a result of these limitations an actual transformer can be considered as ideal only within a limited range of frequencies. (The frequency response of transformers is described in more detail in Chap. XIII.) The third requirement is necessary for high efficiency of power transformation and is introduced into the ideal-transformer definition because it simplifies the expressions for impedance, voltage, and current ratios.

8. Mechanical Analogies in the Impedance-matching Problem.

The impedance of an electric circuit is the complex ratio of voltage applied across two terminals of the circuit to the current at the terminals. In mechanical terms this is analogous to the ratio of force applied to a mechanical system to the velocity of motion of the part of the system where the force is applied, the velocity being assumed in the direction of the force. If the motion is rotary, the mechanical impedance is the ratio of torque to angular velocity. Thus any mechanical device that changes the force-velocity ratio is an impedance-transforming device, or matching system.

A system of gears used to transform the torque-angular-velocity

ratio is a matching device. When an automobile is starting from rest, the power input to the rear wheels must be at high torque and low angular velocity. Later, when the automobile is moving, the required torque is much smaller but the angular velocity is much greater. At the start the impedance presented by the drive shaft is high compared with the impedance when the automobile is moving at moderate speeds on level ground. The automobile engine on the other hand develops a maximum torque at a particular angular velocity and has a maximum power output at a somewhat higher angular velocity. The familiar torque-speed and output-horsepower-speed characteristic curves for an engine are graphical expressions of the optimum impedance conditions for the engine as a generator of power. The transmission in the automobile is the adjustable matching device between the engine and the wheels. Shifting gears changes the impedance ratio.

Consider the pedaling of a bicycle and suppose the two sprockets to be of the same size, this being unusual but possible in the design of a bicycle. Under these conditions the force required to pedal the bicycle will be quite small, for operation on level ground, but it would require very rapid pedaling to get normal speed. The impedance is too low for effective energy transfer to the bicycle. However, this sprocket arrangement would be ideal for riding up steep hills, where the ordinary sprocket ratio would present too high an impedance to the rider so that he would be unable to deliver energy to the pedals at anything like the usual rate.

The handle on an old-fashioned water pump is a matching device to transform the impedance of the piston to a somewhat lower value so that the power source, the arm and the hand of the operator, can work at the ideal force-velocity ratio.

Thus the matching problem is by no means confined to electrical engineering but is a familiar and important problem in many other fields.

CHAPTER VI

TRANSIENTS

1. Introduction.—The study of transients in electrical or mechanical systems involves an analysis of the sequence of events immediately following a disturbance of the equilibrium of the system. Thus the angular displacement of a pendulum from its equilibrium position gives rise to a transient series of oscillations, which eventually dies out, leaving the pendulum in its original position. The same pendulum under water would make fewer

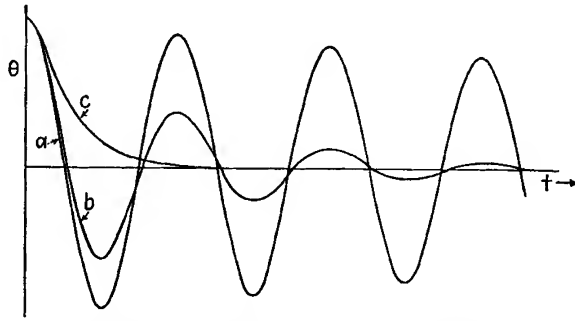


FIG. 1.1.—Transient angular displacement θ of a pendulum plotted against time for three assumed viscosities of the surrounding medium: (a) in air; (b) in water; (c) in heavy oil.

oscillations for the same initial conditions and if in a very viscous oil might not oscillate at all but slowly move back to its position of equilibrium. The three curves in Fig. 1.1 represent the angular displacement of the pendulum for these three transients.

The general objective in the analysis of transients in electrical systems is to express in mathematical form the voltage-time, current-time, and charge-time relationships at various points in the system. Since the analysis of two-element series circuits is simpler than that for the more general case, the transient equations for the resistance-inductance and resistance-capacitance circuits will be developed first.

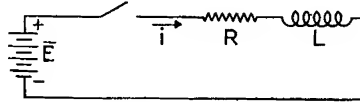
2. Resistance-Inductance Circuit.—Consider the circuit, Fig. 2.1, consisting of a resistance R and an inductance L in series with a

switch and battery. Assume in this and in subsequent circuits that the battery has zero internal impedance. If this is not so, it is necessary merely to add the internal impedance of the battery or generator to the impedance of the circuit and proceed with the analysis. When the switch in Fig. 2.1 is closed, the battery voltage \bar{E} is applied to the circuit and current flows, giving rise to voltage across the resistance and across the inductance. The voltage e_R across the resistor R measured in the direction of the current is

$$e_R = Ri$$

where i is the instantaneous current. The voltage e_L across the inductance L is

$$e_L = -L \frac{di}{dt}$$



where di/dt is the instantaneous time rate of change of the current i .

FIG. 2.1.—Resistance-inductance series circuit.

Note that di/dt in an inductance must always be finite since the voltage e_L across the inductance must always be finite. This means that *the current through an inductance never can jump instantaneously from one value to another*. The change must *always* take a finite time. This is one of the fundamental facts in the analysis of transients.

Conforming to the general convention as to current and voltage signs, currents are positive when they flow clockwise in a given mesh, and voltages also are taken as positive when they are clockwise.

Applying Kirchhoff's voltage law to the circuit of Fig. 2.1,

$$\bar{E} - Ri - L \frac{di}{dt} = 0 \quad (2.1)$$

Rearranging terms to separate the variables,

$$\frac{-R di}{\bar{E} - Ri} = -\frac{R}{L} dt$$

Since a general relationship between current and time is wanted, this equation will be integrated between time limits zero and t , where t is measured from the instant of closure of the switch; since in this case the current must be zero when the switch is open, the current limits of integration are zero and i where i is the instantaneous current at time t .

Introducing these limits,

$$\int_0^i \frac{-R \, di}{\bar{E} - Ri} = -\frac{R}{L} \int_0^t dt$$

Integrating,

$$\log_e \frac{\bar{E} - Ri}{\bar{E}} = -\frac{R}{L} t \quad (2.2)$$

or, in exponential form,

$$i = \frac{\bar{E}}{R} (1 - e^{-Rt/L}) \quad (2.3)$$

The solid line of Fig. 2.2 is a graph of (2.3).

Certain features of this curve are important. The term $e^{-Rt/L}$ approaches zero as t approaches infinity; hence the current

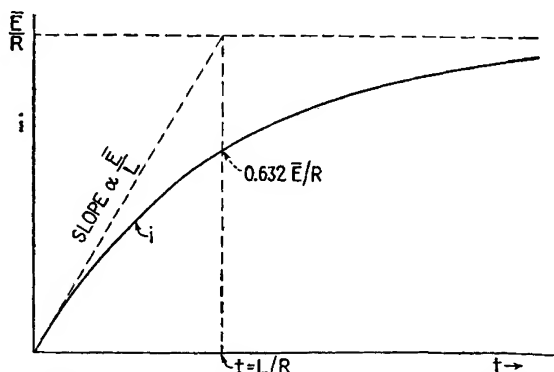


FIG. 2.2.—Graph of current versus time for the series LR circuit of Fig. 2.1.

approaches \bar{E}/R as a final steady value. This final current is shown in the graph as a horizontal dotted line. The voltage drop across the resistance becomes equal to the applied voltage \bar{E} , while the voltage across the inductance becomes zero since the current is constant. On the other hand, at the beginning of the transient the current is zero; hence at the instant of closure of the switch the battery voltage \bar{E} appears as a voltage drop across the inductance.

The slope of the current-time curve is given by the time derivative of (2.3) and is

$$\frac{di}{dt} = \frac{\bar{E}}{L} e^{-Rt/L} \quad (2.4)$$

At time $t = 0$,

$$\left. \frac{di}{dt} \right|_{t=0} = \frac{\bar{E}}{L}$$

This is the initial slope of the current-time curve.

Equation (2.3) indicates that the current never reaches the final value $i = \bar{E}/R$. However, when the value of the exponent Rt/L is numerically equal to 5, the term $\epsilon^{-Rt/L}$ is equal to 0.00674, which subtracted from 1 gives 0.993, so that, when $t = 5L/R$, $i = 0.993\bar{E}/R$. That is, the current has increased from zero to 99.3 per cent of its final value in a time equal to $5L/R$. Since the time for any degree of completion always will be expressed in terms of some numerical constant multiplied by L/R , it is convenient to define the time constant of a resistance-inductance circuit as $t_c = L/R$. At this time,

$$i \Big|_{t=t_c} = \frac{\bar{E}}{R} (1 - \epsilon^{-1}) = 0.632 \frac{\bar{E}}{R} \quad (2.5)$$

In general, the time constant is the time when the *change* in current or voltage from the initial state to the final state is 63.2 per cent complete. The time constant has another interpretation. Suppose that the current kept increasing from zero to its actual final value $i = \bar{E}/R$ at the initial rate of increase $di/dt = \bar{E}/L$. The current would reach its actual final value \bar{E}/R in time L/R , since the tangent of the initial angle is \bar{E}/L . Figure 2.2 illustrates these slopes, currents, and times.

To examine the effects of varying the parameters of the circuit, first let L be increased, everything else remaining unchanged. The effects are a smaller initial slope and a longer time to reach any given percentage of the final value, *i.e.*, a longer time constant. A decrease in L results in a steeper initial slope and a shorter time constant, as indicated in Fig. 2.3. In each case the final value of the current is \bar{E}/R and is the same since R has not been changed.

Varying the magnitude of R does not affect the initial slope but does affect the final current. Increasing R reduces the final current and decreases the time constant. The effects of varying R are shown in Fig. 2.4.

Increasing \bar{E} increases both the initial slope and the final current but does not affect the time constant.

Another type of transient occurs when a current in an inductance dies out. Equation (2.2) is derived by integrating from zero initial current to current i at time t . Since many transients to be dealt

with recur at regular time intervals, it will be necessary usually to integrate from some initial value of current i_1 , where i_1 is the current at the instant the switching or other initiating action occurs.

Figure 2.5 shows a circuit in which the current i for the part of the circuit indicated may flow after the switch is opened. When the switch is closed for the first time, the current through R_1 jumps

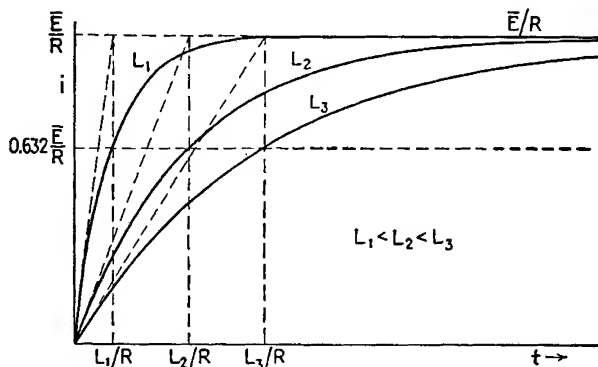


FIG. 2.3.—Effect of varying the inductance L while the resistance R and the applied voltage \bar{E} are held constant.

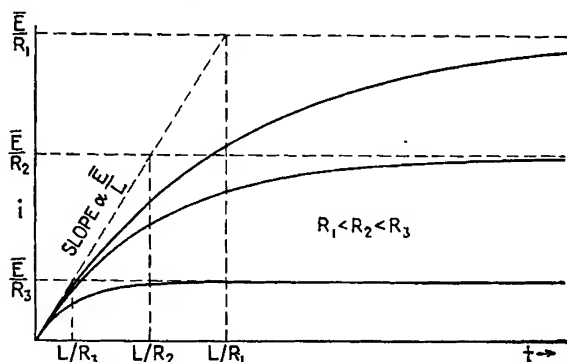


FIG. 2.4.—Effect of varying the resistance R while the inductance L and the applied voltage \bar{E} are held constant.

instantly to the value \bar{E}/R_1 and continues at this value as long as the switch remains closed. The current through R_2 and L is, from (2.3),

$$i = \frac{\bar{E}}{R_2} (1 - e^{-R_2 t/L}) \quad (2.6)$$

Let this current increase from an initial value 0 to a value i_m at t_m . At time t_m let the switch be opened. The voltage equation around

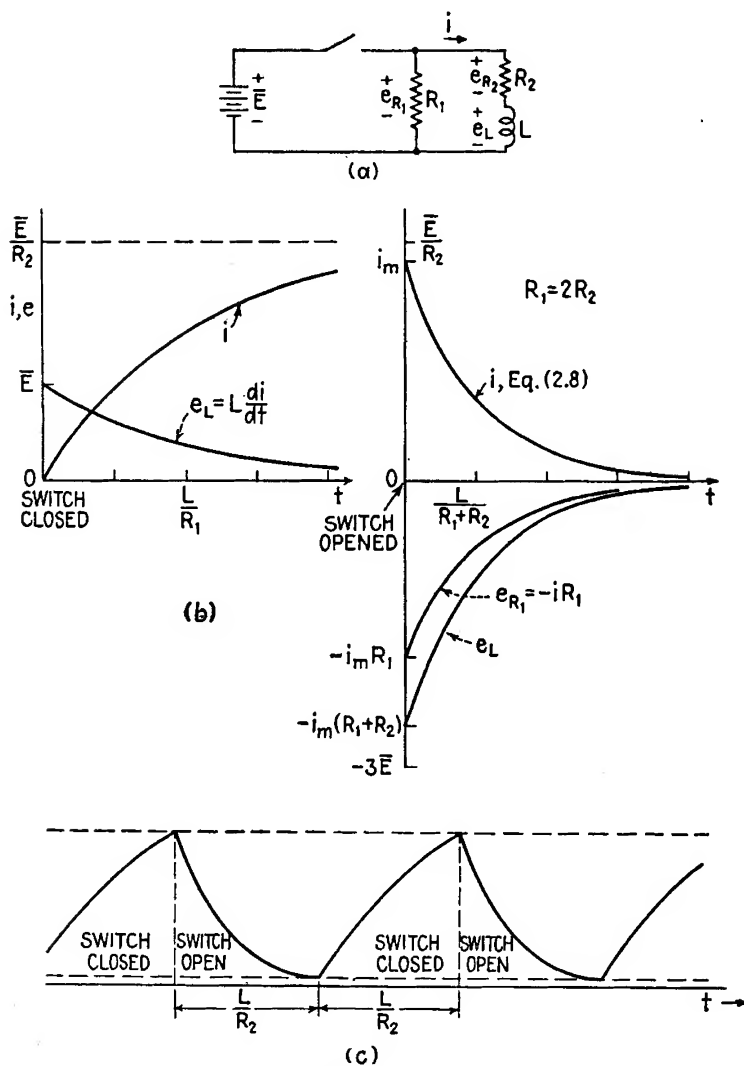


FIG. 2.5.—(a) Circuit containing resistance and inductance; (b) graph of current through R_2 and L when the switch is closed and opened again; (c) graph of current through R_2 and L when switch is closed and opened repeatedly at the interval indicated.

the circuit R_2, L, R_1 is now

$$-i(R_1 + R_2) - L \frac{di}{dt} = 0 \quad (2.7)$$

Since the current through the inductance cannot change instantaneously, the integration is from i_m and the expression for the current is

$$i = i_m e^{-\frac{R_1 + R_2}{L}t} \quad (2.8)$$

Equation (2.8) describes the transient decrease of current in the circuit. Note that now the time constant for the decrease in i is $L/(R_1 + R_2)$, which is smaller than that for the increase. Figure 2.5b shows the transients for current through R_2 and L after closing and also after opening the switch, it being assumed that a long time elapses between the closing and the opening.

If the switch is reclosed, (2.1) applies. If the reclosure of the switch occurs before the current dies out, (2.1) must be integrated from the value of the current i_n at the instant of reclosure of the switch, and

$$i = \frac{\bar{E}}{R_2} (1 - e^{-R_2 t/L}) + i_n e^{-R_2 t/L} \quad (2.9)$$

Figure 2.5c shows the current through L and R_2 resulting from opening and closing the switch at regular intervals. As shown, the interval between opening and closing is equal to L/R_2 , the process being repeated indefinitely. It is assumed for Fig. 2.5 that $R_1 = 2R_2$.

The voltage across R_1 is \bar{E} while the switch is closed. While the switch is open, the voltage is

$$e_{R_1} = -iR_1 = i_m R_1 e^{-\frac{R_1 + R_2}{L}t} \quad (2.10)$$

The maximum value that i_m can have is \bar{E}/R_2 . Then the maximum value of $|iR_1|$ while the switch is open is

$$|e_{R_1}|_{\max} = \bar{E} \frac{R_1}{R_2} \quad (2.11)$$

If R_1 is larger than R_2 , the voltage across R_1 at the time of opening the switch is larger than the battery voltage, as shown in Fig. 2.5 for $R_1 = 2R_2$.

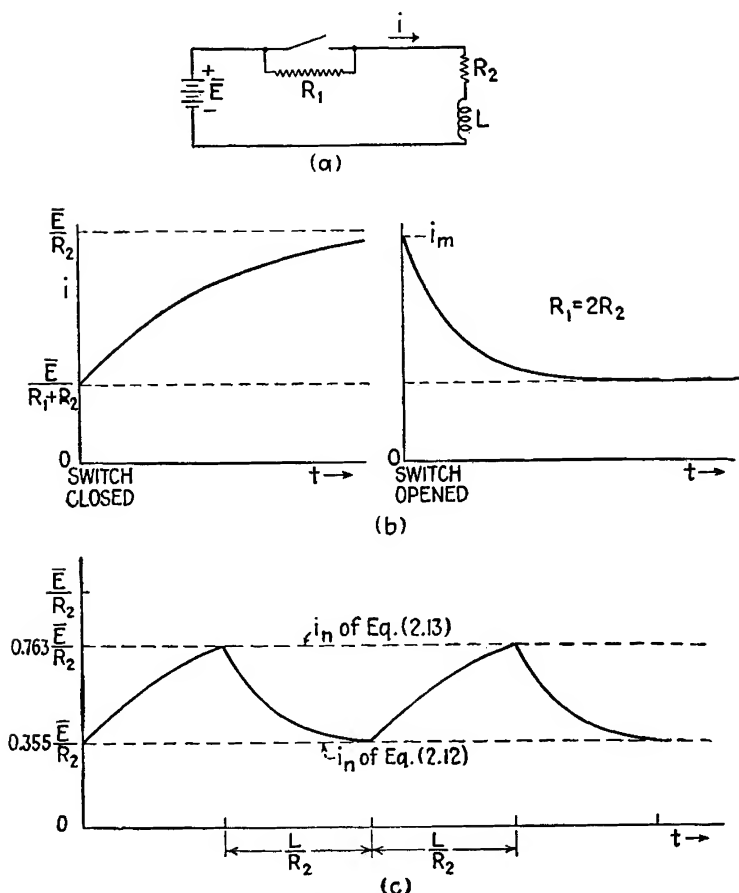


FIG. 2.6.—(a) Circuit containing resistance and inductance; (b) graph of current through R_2 and L when the switch is closed and opened again; (c) graph of current through R_2 and L when the switch is closed and opened repeatedly at the interval indicated.

Figure 2.6a shows a circuit whose transients are similar to those of the previous circuit. When the switch is closed, the current through L is

$$i = \frac{\bar{E}}{R_2} (1 - e^{-R_2 t/L}) + i_n e^{-R_2 t/L} \quad (2.12)$$

where i_n is the current flowing through R_2 and L at $t = 0$, the instant of closure of the switch. When the switch is opened, the current through L is

$$i = \frac{\bar{E}}{R_1 + R_2} (1 - e^{-\frac{R_1 + R_2}{L} t}) + i_m e^{-\frac{R_1 + R_2}{L} t} \quad (2.13)$$

where i_m is the current through R_2 when the switch is opened. Figure 2.6*b* shows the current through R_2 as a function of time after closing and also after opening the switch, a long time elapsing after opening the switch so that the steady state exists. Figure 2.6*c* shows the recurring transients that occur when the switching is repeated at regular intervals, the interval between operations being L/R_2 . Again in this circuit there will be a voltage developed across R_1 when the switch is opened. At the instant of opening the switch the voltage across R_1 is greatest and is

$$e_{R_1} = i_m R_1$$

which can have a maximum value

$$|e_{R_1}|_{\max} = \bar{E} \frac{R_1}{R_2} \quad (2.14)$$

Here again the voltage across R_1 at the instant of opening the switch may be many times the battery voltage.

An extreme case of this type of transient arises when the switch in a circuit like that of Fig. 2.1 is opened. In this case the resistance in parallel with the switch is infinite; and if there were no other limiting factors, the voltage across the switch would be infinite when the switch is opened, according to (2.14), where R_1 is infinite. One of the limiting factors is the capacitance between the switch contacts as they part. This effect will be discussed later. Also as a result of the high voltage developed across the switch when opened, an arc strikes across the contacts so that opening the switch inserts, not an infinite resistance in series with the circuit, but the finite resistance of the arc. Since the resistance of an arc is not fixed, it is difficult to analyze the problem and calculate the exact voltage developed across the opening contacts. There is, however, the possibility of developing high voltages for brief periods of time by opening inductive circuits. These voltages may be destructive since they may cause insulation breakdown in the inductance or elsewhere in the circuit. The arc at the switch causes pitting and burning of the contact surfaces. A circuit for the suppression of such arcs will be discussed later.

Equations involving two exponential terms, such as (2.9), (2.12), (2.13), may be thought of as representing two simultaneous transients in the circuit: the building up of the current to the new steady state as expressed by the first term, and the dying out of the current from the initial value as expressed by the second term.

Figure 2.7 shows the two terms of (2.12) plotted separately and in sum in order to bring out the further significance of the time constant and the initial slopes. Note that the change from the initial value i_n to the final value \bar{E}/R_2 is 63.2 per cent complete at the time $t_c = L/R_2$. The curve for i is the same as would be obtained if the curve $(\bar{E}/R_2)(1 - e^{-R_2 t/L})$ were moved bodily to the left until it intersected the vertical axis at i_n and only the part to the right of the vertical axis were retained.

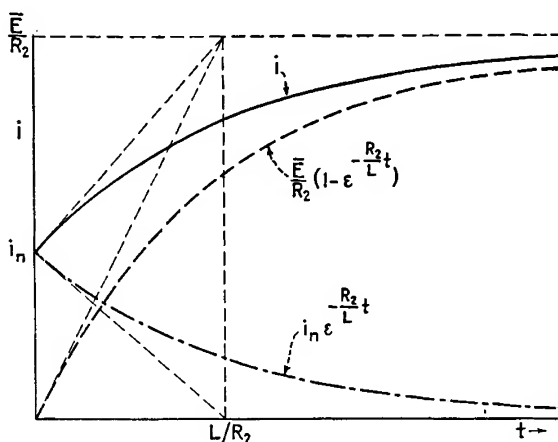


FIG. 2.7.—Graph of equation (2.12) showing the terms plotted separately and in total.

Consider the energy input by the battery to the LR circuit of Fig. 2.1. At any instant the voltage across the inductance is $L di/dt$, and the current is i . Therefore the instantaneous power input to the inductance is $p = e_L i_L = Li di/dt$ watts, and the total energy input in watt-seconds is

$$U_L = \int_0^t Li \frac{di}{dt} dt$$

or in terms of current

$$U_L = \int_0^i Li di = \frac{1}{2} Li^2 \quad (2.15)$$

This is the total energy input to the inductance when the current is increased from zero to i . The energy input to an inductance is stored energy because it is released into the circuit when the current is reduced to zero again. Part of this energy is dissipated in the arc across the switch contacts when an inductive circuit is opened. The

energy stored in an inductance is in the form of a magnetic field around the conductors that form the inductance. An increase in the current through a given conductor causes the number of magnetic lines of force in the field to increase, and it is the cutting of the conductor by these lines of force that sets up the voltage $L di/dt$ that opposes the increase of current. The emf applied to the circuit does work in establishing the current against this counter emf $= L di/dt$. Any attempt to diminish the current brings about the collapse of the magnetic field, which again cuts the conductor; this time the magnetic field is changed in the opposite direction and induces a voltage that tends to maintain the current. The action of the inductance is to oppose any change in current.

In mechanics, the quantity mass is analogous to inductance.

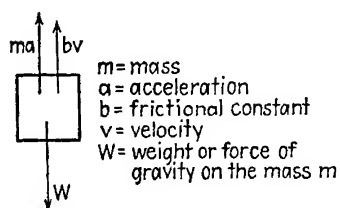


FIG. 2.8.—Free-body diagram of a body falling in a resisting medium under a constant downward force W , its weight.

Consider the forces acting on a mass falling in a uniform gravitational field. If all the forces that act on a falling body are indicated as vectors, the result is the so-called “free-body” diagram, Fig. 2.8. The drawing of the free-body diagram of a single particle is the mechanical equivalent of indicating the various voltages around a single closed electric circuit. Equating to zero the

sum of all the forces on the falling body,

$$W - ma - bv = 0$$

or

$$m \frac{dv}{dt} + bv = W \quad (2.16)$$

Integrating and solving for v ,

$$v = \frac{W}{b} (1 - e^{-bt/m}) \quad (2.17)$$

or if there is an initial downward velocity v_0 ,

$$v = \frac{W}{b} (1 - e^{-bt/m}) + v_0 e^{-bt/m} \quad (2.18)$$

These equations for velocity have the same form as (2.3) and (2.9) for current in resistance-inductance circuits. If mass is considered as the mechanical analogue of inductance, velocity is analogous to

current and force to voltage. The quantity b in (2.16) is a sort of mechanical resistance that depends on the viscosity of the medium in which the body falls and on the shape of the body. Thus a man falling through the air with his parachute unopened would have a relatively small value of b , his final velocity would be quite large, and the time required to reach any proportion of this final velocity would be relatively long. The man falling with the parachute open would have a much smaller final velocity and would reach that velocity in a relatively shorter time, this being an example of a large value of b and consequently a small value of the time constant m/b . If the man could alternately open and close the parachute at regular intervals, the graph of his downward velocity vs. time would have the same form as that of Fig. 2.6c.

In connection with Figs. 2.5, 2.6, attention was called to the possibility of producing large voltages from relatively small ones by means of the circuits shown. A mechanical analogue of this condition is force developed by a hammer upon striking a nail. It is impossible for a man to exert sufficient force with his hand to push a spike into a piece of lumber. However, by exerting a relatively small force over a period of time while swinging the hammer, he can accelerate the mass of the hammer up to a certain velocity. When this mass strikes the head of the spike, the stored kinetic energy is given up to the spike, the forward velocity of the hammer is abruptly reduced, and the resultant force is sufficient to push the nail into the wood. The operation of the hydraulic ram is another good example of this sort of mechanical transient.

The work done in accelerating a mass is given by the familiar work integral

$$\begin{aligned} U &= \int \text{force} \cdot \text{distance} \\ &= \int F \, ds = \int ma \, ds \\ &= \int m \frac{dv}{dt} \, ds \end{aligned}$$

but since

$$\begin{aligned} v &= \frac{ds}{dt} \\ U &= \int_0^v mv \, dv = \frac{1}{2}mv^2 \end{aligned} \tag{2.19}$$

This is the expression for the kinetic energy of a mass in motion and is analogous to the magnetic energy stored in an inductance

$$U_L = \frac{1}{2}Li^2$$

3. Capacitance-Resistance Circuit.—The range of possible time constants for an inductance-resistance combination is restricted

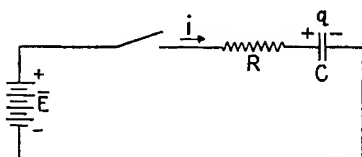


FIG. 3.1.—Circuit of resistance and capacitance in series with a switch and battery.

somewhat by the fact that a physical inductor has an appreciable inherent resistance. Thus the time constant for an inductance-resistance circuit cannot be greater than L/R_L where L and R_L are the inductance and resistance of the inductor. There is no such

limitation in capacitance-resistance circuits since it is possible to build capacitors in which the series resistance is negligibly small. The capacitance-resistance circuit has many and varied applications in electronic equipment.

Figure 3.1 shows a circuit made up of resistance and capacitance in series with a switch and a battery. Kirchhoff's voltage equation for the circuit is

$$\bar{E} - Ri - \frac{q}{C} = 0 \quad (3.1)$$

where \bar{E} is the battery voltage, $-Ri$ is the voltage across the resistance, and $-q/C$ is the voltage across the capacitance. The current i is assumed positive when it flows clockwise in the circuit, and the charge q on the capacitor is assumed positive when the voltage across the capacitor is positive on the left as indicated. The current into a capacitor equals the rate of increase of charge, $i = dq/dt$. Since the current cannot be infinite, the charge on the capacitor cannot change at an infinite rate. Therefore, *the voltage across a capacitor cannot jump suddenly from one value to another*. This is another fundamental concept in the study of transients.

Equation (3.1) may be written in terms of q as

$$\bar{E} = R \frac{dq}{dt} + \frac{q}{C} \quad (3.2)$$

This is of the same form as (2.1). Integration between the limits of initial charge, q_0 at time $t = 0$, to a general value of q at any time t results in

$$q = C\bar{E}(1 - e^{-t/CR}) + q_0 e^{-t/CR} \quad (3.3)$$

The corresponding current is

$$i = \frac{dq}{dt} = \frac{\bar{E} - \frac{q_0}{C}}{R} e^{-t/CR} \quad (3.4)$$

Figure 3.2 is a plot of (3.3) where the initial charge on the capacitor is zero, that is, $q_0 = 0$. The graph shows the relationship between

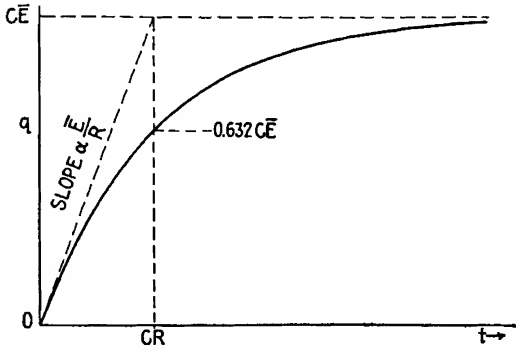


FIG. 3.2.—Graph of charge vs. time for the circuit of Fig. 3.1. when the initial charge on the capacitor is zero.

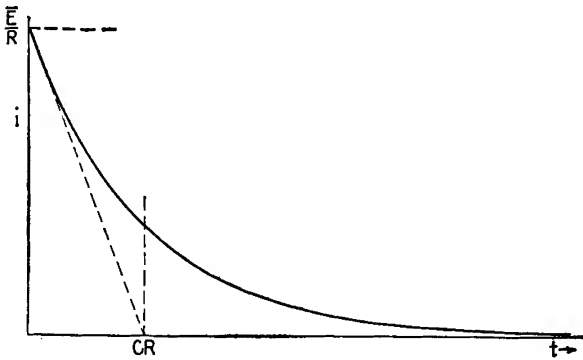
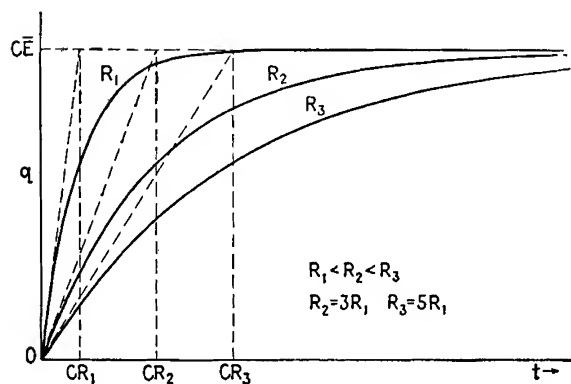
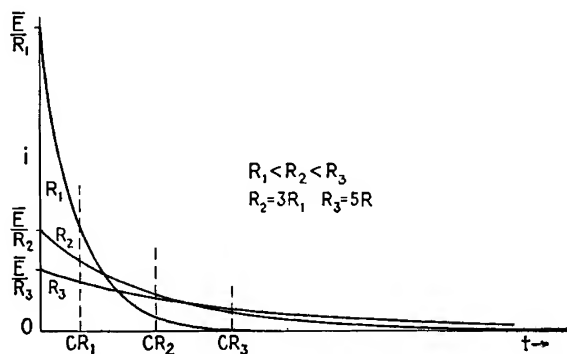
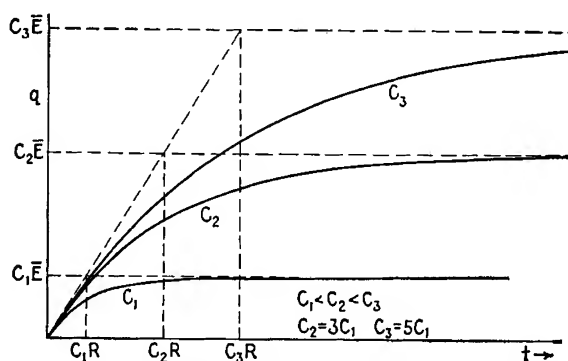


FIG. 3.3.—Graph of current vs. time for the same conditions as those of Fig. 3.2.

the time constant RC , the initial slope \bar{E}/R , and the final value of the charge $C\bar{E}$. Figure 3.3 is the graph of (3.4) showing the current for the same transient.

Figures 3.4 and 3.5 show the effects of varying the series resistance R . Since the final value of charge (after an infinite time) is independent of the resistance, all the charge curves, Fig. 3.4, approach the same final value at rates depending on the time constant. The corresponding current curves, Fig. 3.5, all enclose the

FIG. 3.4.—Graph of charge vs. time, showing the effect of varying the resistance R .FIG. 3.5.—Graph of current vs. time, showing the effect of varying the resistance R .FIG. 3.6.—Graph of charge vs. time, showing the effect of varying the capacitance C .

same area between zero and infinity since the area under the curve is $\int_0^\infty i \, dt = C\bar{E}$, which is constant when C and \bar{E} are constant.

Figures 3.6 and 3.7 show the results of varying the capacitance. When an external voltage is suddenly applied, it must all appear across the resistance, since the voltage across a capacitance cannot change instantaneously. Thus the initial current is \bar{E}/R and is the same for all values of C .

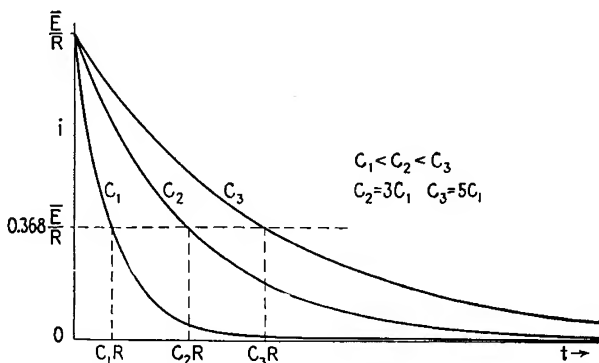


FIG. 3.7.—Graph of current vs. time, showing the effect of varying the capacitance C .

If the charge on the capacitor is not zero when the transient is initiated, the expression for the charge is

$$q = C\bar{E}(1 - e^{-t/CR}) + q_0 e^{-t/CR} \quad (3.5)$$

where q_0 is the initial charge on the capacitor, and the corresponding current equation is

$$i = \frac{\bar{E}}{R} e^{-t/CR} - \frac{q_0}{CR} e^{-t/CR} = \frac{\left(\bar{E} - \frac{q_0}{C}\right)}{R} e^{-t/CR} \quad (3.6)$$

Notice that the quantity $(\bar{E} - q_0/C)$ is the algebraic sum of the battery voltage and the initial voltage across the capacitor. Equations (3.5), (3.6) are plotted in Figs. 3.8, 3.9.

With initial charge on the capacitor the time constant has the same significance as formerly, being the time required for the charge from the initial to the final value of charge or current to be 63.2 per cent complete.

The first term on the right of (3.5) represents a transient in which the capacitor is charged from zero charge to final charge $C\bar{E}$.

This term is plotted as a dash line in Fig. 3.8. The second term on the right of (3.5) represents the discharge through the circuit of the initial charge q_0 . This term is plotted as a dot-dash line in Fig. 3.8. Thus the actual charge of the capacitor, shown as a solid line in Fig. 3.8, may be regarded as the superposition of two transient

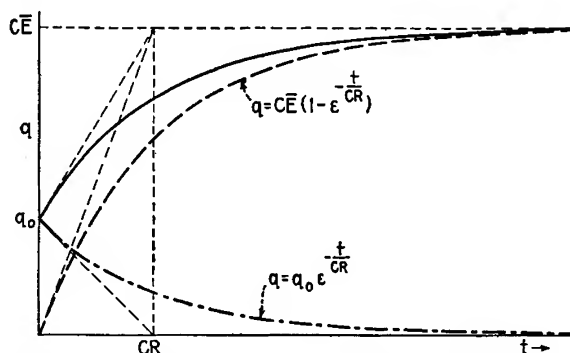


FIG. 3.8.—Graph of equation (3.5) when the initial charge on the capacitor is positive.

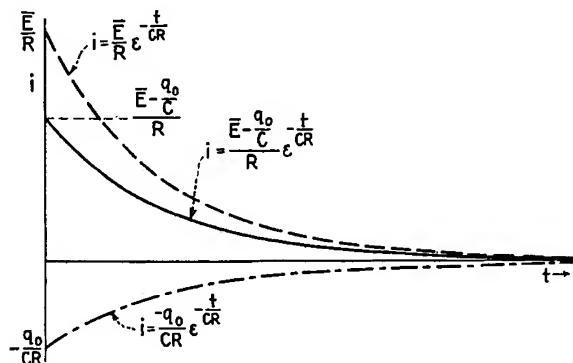


FIG. 3.9.—Graph of equation (3.6) when the initial charge on the capacitor is positive.

effects, one the building up of the new charge and the other the decay of the old charge.

The solid line of Fig. 3.8 is a section of the dash curve; if the dash curve is moved to the left until it intersects the vertical axis at q_0 , the part to the right of the vertical axis coincides with the solid-line curve. This is a particular example of a general principle that the current-time and charge-time curves for any initial condition are sections of the curves obtained when the capacitor has no

initial charge. Figure 3.9 shows the actual current (solid line) as the sum of two transient currents.

As an illustration of this principle, suppose the switch of Fig. 3.1 is closed and then reopened before the charging operation has been completed, then closed again and reopened, this process being repeated several times. During each interval when the switch is closed, the capacitor accumulates charge. During each interval

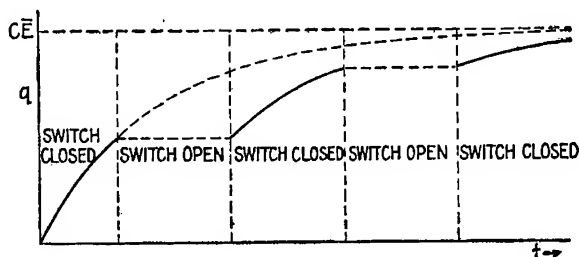


FIG. 3.10.—Effect of closing and opening the switch of Fig. 3.1 several times; graph of charge vs. time.

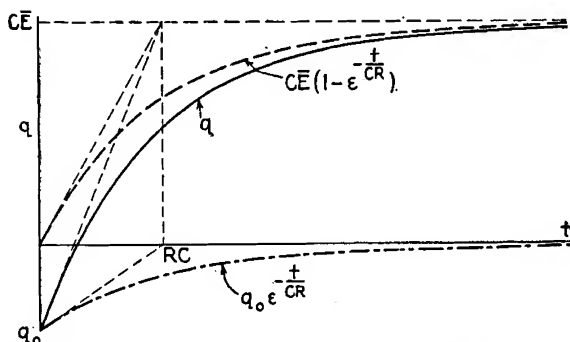


FIG. 3.11.—Graph of equation (3.5) when the initial charge on the capacitor is negative.

when the switch is open, the charge remains fixed. The resulting charge vs. time curve is shown in Fig. 3.10. The rising portions of this curve are portions of the curve that would be obtained if the switch had remained closed. That is, if the intervals when the switch is opened were disregarded and the rising portions of the curve shifted horizontally to the left to eliminate the horizontal portions, the resultant curve would be the same as though the switch had never been opened.

The solid-line curves of Figs. 3.11 and 3.12 are graphs of (3.5) and (3.6) when the initial charge of the capacitor is negative, the

dash and dot-dash curves representing the first and second terms on the right of (3.5) and (3.6) for this case. The dash curves would obtain if the initial charge were zero, and the curves for the actual charge and current are sections of the dash curves moved to the right.

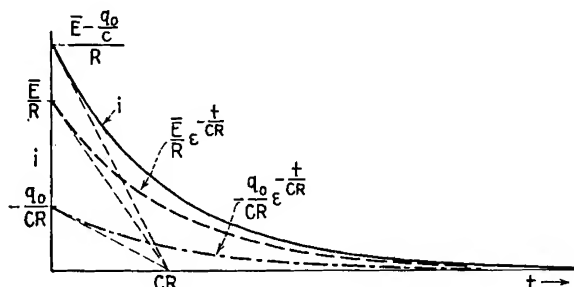


FIG. 3.12.—Graph of equation (3.6) when the initial charge on the capacitor is negative.

The case of negative initial charge is of particular importance when a square-wave voltage is applied to the capacitance-resistance circuit.

In a series capacitance-resistance circuit, the current approaches zero after a long time, and the charge q on the capacitor approaches $C\bar{E}$.



FIG. 3.13.—Simple form of series CR circuit.

Thus the voltage across the capacitor becomes equal and opposite to the battery voltage. If now the switch is opened and the battery is replaced by a wire, reclosing the switch will initiate another transient. The

circuit is shown in Fig. 3.13. The sum of the voltages around the circuit is

$$iR + \frac{q}{C} = 0 \quad (3.7)$$

so that

$$R \frac{dq}{dt} = -\frac{q}{C}$$

and

$$\frac{dq}{q} = -\frac{1}{RC} dt$$

Integrating from $q = q_0$ (the initial charge at the time of closing the switch),

$$q = q_0 e^{-t/CR} \quad (3.8)$$

and

$$i = -\frac{q_0}{CR} e^{-t/CR} \quad (3.9)$$

Graphs of the complete charge and discharge are shown in Fig. 3.14. The current during discharge is opposite in direction to that during charge and is plotted negatively. The same time constant $t_c = CR$ holds for all the curves and is a function of C and R only and not of initial or boundary conditions.

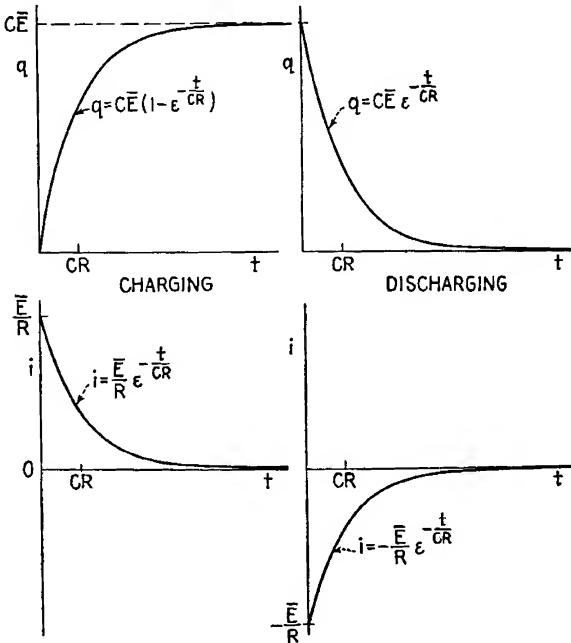


FIG. 3.14.—Curves of charge and current for a charging operation and a discharging operation.

The quantity mass in mechanics was shown to be analogous to inductance in an electric circuit. Capacitance also has an analogue in mechanics, expressible in terms of the elastic deformation of mechanical systems under mechanical forces. The compressing of a spring or a volume of gas is analogous to the charging of a capacitor. Hooke's law relates the displacement to the force producing the displacement and to the stiffness of the spring or other elastic body.

4. Applications of CR Circuits.— CR circuits having different time constants are extremely useful in electronic applications.

The characteristics of these circuits are brought out by considering their response to an applied voltage having square or rectangular waveform. Such a voltage is usually generated by an electronic device, but the result is equivalent to a battery applied to a circuit through a quick-reversing commutator in such a manner that there

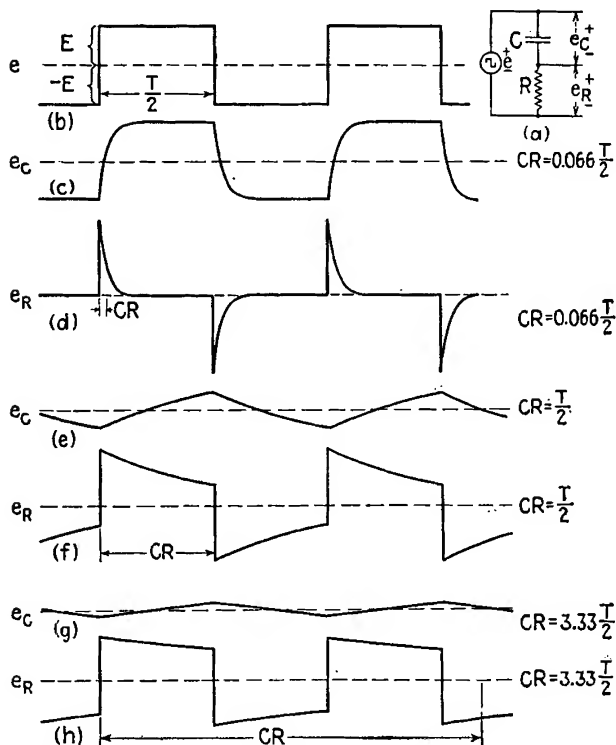


FIG. 4.1.—Voltage waveforms resulting from the application of a square wave to a CR circuit, for the time constants indicated.

is no interval when the circuit is open and the applied voltage has equal intervals of positive and negative polarity.

Figure 4.1 shows the voltage across the elements of a series CR circuit having various time constants, the applied voltage having the square-wave form of Fig. 4.1b. The curves c and d are the voltage across the capacitor and the voltage across the resistor in a circuit whose time constant is very small compared with the period T of the square wave. The curves e and f are for a circuit whose time constant is equal to the half period $T/2$ of the square wave. The curves g and h are for a time constant larger than T .

The circuit having a very small time constant is often called a differentiator or peaker circuit. Curve d shows that the voltage across the resistance is a series of sharp peaks, whence the name peaker circuit. Since the voltage across the resistance is iR , the voltage consists of a series of peaks each having the form of Fig. 3.3. The small time constant leads to a rapid completion of the transient so that the duration of the peaks, or pulses, is very short. As shown in Fig. 4.1c, the capacitor C is charged to $+E$ practically at the end of a positive half cycle and to $-E$ at the end of a negative half cycle. At the beginning of each half cycle of the square wave, the applied voltage and the voltage previously built up across C act in the same direction. Therefore the peaks of curve d have a value approximately equal to $2E$. This is an example of the principle enunciated in Sec. 3, that a suddenly applied voltage appears entirely across the resistance in a CR circuit, since the voltage across a capacitance cannot change suddenly. The alternations of the square wave involve a sudden change of $2E$ volts (from $\pm E$ to $\mp E$). All this sudden change appears across R .

The name differentiator for this circuit arises from a consideration of (3.1) when the applied voltage is a function of time $e(t)$. Then

$$e(t) = Ri + \frac{q}{C} \quad (4.1)$$

By making R and C small, the Ri term can be made small compared with q/C , so that

$$q \doteq Ce(t)$$

and therefore

$$i = \frac{dq}{dt} \doteq C \frac{de(t)}{dt}$$

The voltage across the resistance is then

$$e_R = Ri \doteq RC \frac{de(t)}{dt} \quad (4.2)$$

Thus e_R is approximately proportional to the derivative of the applied voltage.

Note that the approximation used requires that

$$\frac{q}{C} \gg Ri \quad \text{or} \quad \frac{q}{CR} \gg \frac{dq}{dt}$$

When q is zero, this condition is violated. Furthermore, if dq/dt is large as it is when de/dt is large, it is impossible to maintain the condition. Whereas the value of $de(t)/dt$ is infinite at the instances of reversal of the square wave and zero at other times, the output voltage of the differentiator circuit is not infinite but is equal to $2E$ at the instants of reversal and is not zero at other times.

However, by making CR very small compared with the duration of the cycle it is possible to approach the conditions of (4.2) when the derivative is finite. Making CR small also makes the output voltage small.

The curves e and f for the circuit having a time constant equal to $T/2$ show that the reversal in the voltage across C is considerably delayed with respect to the instants at which the applied voltage reverses. This circuit is used in certain applications where the delay is the desired feature.

The circuit having a very large time constant is employed commonly in the form of the coupling capacitor and grid resistor in a resistance-coupled amplifier. In this application the voltage across the resistor should be the same as the voltage input. To secure this result, the Ri term in (4.1) should be much greater than the q/C term in order that

$$e_R = Ri \doteq e(t) \quad (4.3)$$

With this end in view it is necessary to make both C and R large or the time constant CR large compared with the period of the input voltage. Thus Fig. 4.1*h* is a close approximation to the original square wave.

In some applications of the CR circuit having a large time constant, the output voltage is taken from the capacitor and the circuit is called an integrator circuit. Thus, the curve of Fig. 4.1*g* has the same form as the integral of the function of Fig. 4.1*b*. This integrator action results when CR is very large compared with the period of the square wave, since under these conditions, the capacitor accumulates only a small charge; the q/C term in (3.1) is then small compared with the Ri term. Suppose any voltage waveform $e(t)$ is applied to a CR circuit so that $e(t) = q/C + Ri$. If q is kept small and the product CR is large, which is true when the voltage across the capacitor is much less than that across the resistor,

$$e(t) \doteq Ri \quad \text{or} \quad i \doteq \frac{e(t)}{R}$$

and since $q = \int i \, dt$ the voltage across the capacitor, except for a constant of integration, will be given by

$$e_c = \frac{q}{C} \doteq \frac{1}{CR} \int e(t) \, dt \quad (4.4)$$

When variational components are of interest, the constant term may be neglected.

Thus the variational part of the output voltage is approximately proportional to the time integral of the variational part of the applied voltage. Figure 4.1b shows an applied voltage square wave and Fig. 4.1g shows the output waveform of the integrator circuit when the time constant of the integrator circuit is 1.67 times the period T of the square wave.

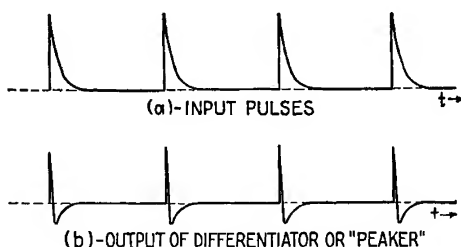


FIG. 4.2.—Use of a CR peaker circuit for sharpening pulses.

Integrator and differentiator circuits are also possible with resistances and inductances. The differentiator, or peaker, circuit is commonly used as a pulse-sharpening circuit. For example, the rather broad pulses of Fig. 4.2a when applied to a differentiator circuit whose time constant CR is small compared with the duration of each pulse yield the extremely sharp pulses of Fig. 4.2b. The negative portion of the voltage of Fig. 4.2b can be removed by methods to be discussed in Chap. XXIV and the sharp positive pulses retained.

As an example of the use of an integrator consider the simple relaxation oscillator shown in Fig. 4.3. The thyatron (type 884 tube) in the circuit is biased to ignite at some voltage less than the applied battery voltage \bar{E} . The resistance R_2 is much less than R_1 . When the switch is closed, a current through R_1 charges the capacitor C . No current flows through R_2 since the thyatron does not conduct until a certain ignition voltage is reached, determined by and approximately proportional to the grid-bias voltage. Thus the capacitor charges until its voltage reaches the "firing"

value for the tube. It is the property of gas tubes such as this one that, once conduction has started, the voltage drops to a low value (about 15 volts) which is nearly independent of the current through the tube. Since R_2 is small and the tube drop is small, the time constant during the discharge of the capacitor through R_2 and the tube is much less than that during the charging of the capacitor through R_1 . When the current through the tube falls to a very low value, the tube stops conducting and the capacitor is recharged through R_1 as before. This process repeats itself in a

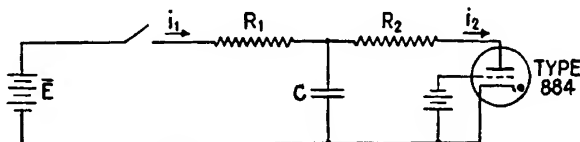


FIG. 4.3.—Simple relaxation oscillator.

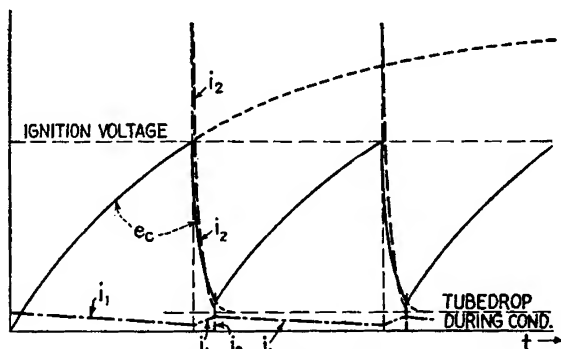


FIG. 4.4.—For the circuit of Fig. 4.3, graphs versus time of voltage e_c across the capacitor (solid line), current i_2 through R_2 (dash line), and current i_1 through R_1 (dot-dash line).

regular cycle. The whole combination of circuit and tube is known as a relaxation oscillator. Figure 4.4 shows the voltage e_c across the capacitor as a function of time. The values of ignition voltage and extinction voltage of the thyatron form the upper and lower boundaries of the voltage variation. If the constants are chosen so that (1) the charging curve of the capacitor is approximately a straight line, (2) the discharge time is much shorter than the charge time, the result is a saw-tooth wave that finds practical application as the sweep voltage for a cathode-ray oscilloscope. Condition (1) is necessary in order to make the horizontal traverse of the spot on the screen linear with respect to time. This is accomplished by having the applied d-c potential large compared with the firing voltage of the tube, the operation being thus con-

fined to a small portion of the total charging curve. Another more elaborate means of achieving this result is discussed in Chap. XXIV. Condition (2) is desirable in order to make the return trace on the oscillograph screen occupy as short a time as possible. This is accomplished by making CR_2 as small as possible. It might be supposed that R_2 could be made zero, and this is actually done in some cases. The real limitation in this direction is that of limiting the peak current in the thyratron in order to avoid damage to the emitter. If voltages are not too high and if C is small, it is quite possible that wiring resistance and inductance plus tube drop will hold the current to a safe value. The rate of repetition of the cycle, *i.e.*, the sweep frequency, may be controlled by varying the time constant CR_1 . Since R_1 must be kept much larger than R_2 , the range of variation is somewhat limited. Furthermore, R_1 must

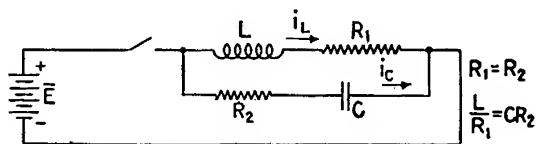


FIG. 4.5.—Spark-suppressor circuit.

be so large that at the end of the discharge of the capacitor the battery cannot maintain enough current in the tube to maintain the arc. In the usual application the variation of a part of R_1 is used as a vernier frequency control while the coarse control of frequency is accomplished by changing the value of C in steps.

Figure 4.4 also shows the currents through R_1 and R_2 during the cycle. The voltage across R_2 has the same waveform as the current and is a series of pulses. The relaxation oscillator here described is also used as a pulse generator.

Another application of CR circuits is that of suppression of the transient in an LR circuit. It is pointed out in Sec. 2 that the sudden opening of an LR circuit may result in dangerously high voltages and burning of contact points. This effect may be minimized by the use of the circuit shown in Fig. 4.5. With $R_1 = R_2 = R$ and with the time constants of the two branches of the circuit equal, that is, $L/R_1 = CR_2$ or $L/C = R^2$, the two currents after closing the switch are

$$i_L = \frac{\bar{E}}{R_1} (1 - e^{-R_1 t/L}) \quad (4.5)$$

$$i_C = \frac{\bar{E}}{R_2} e^{-t/RC} \quad (4.6)$$

and the total current through the switch is the sum of these,

$$i = i_L + i_C = \frac{\bar{E}}{R} \quad (4.7)$$

Thus current in the switch rises immediately to its final value and remains there. The transient components cancel exactly. If now the switch is opened, the current through the switch must be zero so that

$$i = i_L + i_C = 0$$

and under the conditions assumed,

$$i_L = \frac{\bar{E}}{R} e^{-Rt/L}$$

$$i_C = -\frac{\bar{E}}{R} e^{-t/RC}$$

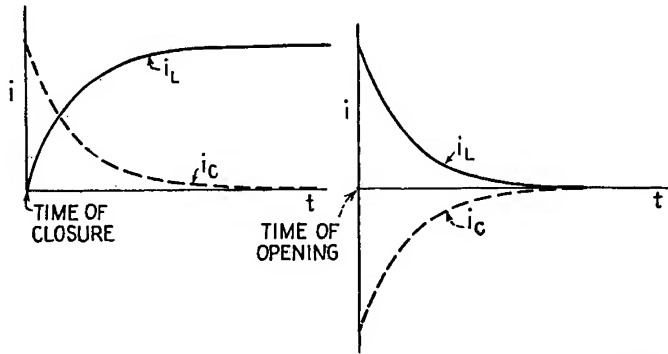


FIG. 4.6.—Graphs of current in the two branches of the circuit of Fig. 4.5.

Then the voltage across the LR circuit is

$$Ri + L \frac{di}{dt} = \bar{E}e^{-Rt/L} - \bar{E}e^{-Rt/L} = 0$$

Hence the voltage across the switch never exceeds the battery voltage \bar{E} . Figure 4.6 shows the currents in each branch after closing and after opening the switch. This circuit finds practical application in the suppression of arcing at key and relay contact points. In an automobile ignition system the breaker points are by-passed by a capacitor to achieve the same result. In this application, R_2 of Fig. 4.5 is part of R_1 , which is the resistance of the primary of the ignition coil, and C is connected across the contacts of the breaker. The transient analysis is similar. The

voltage across C and hence across the contacts never exceeds the battery voltage in a circuit adjusted to critical damping, Sec. 5.

It is also a property of the circuit of Fig. 4.5 that its impedance is a pure resistance equal to R_1 or $R_2 = \sqrt{L/C}$ at all frequencies.

5. Series LCR Circuit.—Consider the series circuit containing inductance, capacitance, resistance, connected to a battery and switch, Fig. 5.1. Kirchhoff's voltage law applied to this circuit gives

$$\bar{E} - Ri - L \frac{di}{dt} - \frac{q}{C} = 0 \quad (5.1)$$

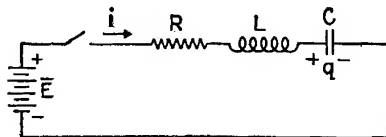


FIG. 5.1.—Series LCR circuit.

This equation might be written either in terms of the charge q on the capacitor or in terms of the current i . Since $i = dq/dt$, (5.1) becomes

$$L \frac{d^2q}{dt^2} + R \frac{dq}{dt} + \frac{q}{C} = \bar{E} \quad (5.2)$$

Differentiating (5.1),

$$L \frac{d^2i}{dt^2} + R \frac{di}{dt} + \frac{i}{C} = 0 \quad (5.3)$$

Equation (5.3) is homogeneous; hence its solution is somewhat simpler than that of (5.2). As the solutions of the LR and CR circuits were exponential functions of time, it is reasonable to suppose that the solution of (5.3) is exponential. If the current i is assumed to have the form $i = Ae^{kt}$, the first and second derivatives will be

$$\frac{di}{dt} = Ak e^{kt} \quad \frac{d^2i}{dt^2} = Ak^2 e^{kt}$$

Substituting these values in (5.3),

$$Lk^2 + Rk + \frac{1}{C} = 0 \quad (5.4)$$

Solving for k ,

$$k = -\frac{R}{2L} \pm \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}} \quad (5.5)$$

The constant k has two values corresponding to the plus and minus sign in front of the radical. It is therefore necessary to take account of both signs in the expression for the current. This

can be done best by writing the current in the somewhat more general form

$$i = A_1 e^{k_1 t} + A_2 e^{k_2 t} \quad (5.6)$$

where

$$k_1 = -\frac{R}{2L} + \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}} \quad (5.7)$$

and

$$k_2 = -\frac{R}{2L} - \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}} \quad (5.8)$$

The constants k_1 and k_2 are functions of the circuit elements only, whereas the constants A_1 and A_2 are functions of both circuit elements and the way in which the transient is initiated, that is, A_1 and A_2 are said to depend on the initial conditions of the problem. For the general case it would be necessary to know what voltages were applied and where, as well as the initial charges on all capacitors in the circuit and the initial currents in all inductive elements.

In the circuit of Fig. 5.1 there can be no current when the switch is open. Suppose that the initial charge on the capacitor is q_0 and that the battery voltage \bar{E} is applied by closing the switch. The initial conditions then are

$$t = 0 \quad i = 0 \quad q = q_0$$

the initial current i being zero because the current cannot jump in value after closing the switch owing to the presence of L .

Substituting these values in (5.6),

$$0 = A_1 + A_2$$

whence

$$A_2 = -A_1$$

Upon dropping the subscripts and writing $A = A_1 = -A_2$, the current becomes

$$i = A(e^{k_1 t} - e^{k_2 t}) \quad (5.9)$$

In order to evaluate A consider the voltage equation around the circuit when $t = 0$,

$$\bar{E} - L \left. \frac{di}{dt} \right|_{t=0} - \frac{q_0}{C} = 0$$

so that

$$\left. \frac{di}{dt} \right|_{t=0} = \frac{\bar{E} - \frac{q_0}{C}}{L} \quad (5.10)$$

The initial slope of the current-time curve is given by the derivative of (5.9),

$$\frac{di}{dt} = Ak_1\epsilon^{k_1t} - Ak_2\epsilon^{k_2t}$$

At time $t = 0$,

$$\left.\frac{di}{dt}\right|_{t=0} = A(k_1 - k_2)$$

Substituting the values of k_1 and k_2 ,

$$\left.\frac{di}{dt}\right|_{t=0} = 2A \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}} \quad (5.11)$$

Equating (5.11) and (5.10),

$$A = \frac{\bar{E} - \frac{q_0}{C}}{2L \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}}$$

Thus the equation for the current is

$$i = \frac{\bar{E} - \frac{q_0}{C}}{2L \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}} (\epsilon^{k_1t} - \epsilon^{k_2t}) \quad (5.12)$$

Equation (5.12) can be written in several useful forms according to the relative magnitudes of the quantities $1/LC$ and $R^2/4L^2$ appearing in the radical.

Since the quantity $\sqrt{1/LC - (R^2/4L^2)}$ has the dimensions of frequency, let

$$\omega_0 = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}} \quad (5.13)$$

Multiplying both sides of (5.13) by $j = \sqrt{-1}$,

$$j\omega_0 = \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}$$

Substituting this in (5.12),

$$i = \frac{\bar{E} - \frac{q_0}{C}}{2jL\omega_0} \epsilon^{-Rt/2L} (\epsilon^{j\omega_0t} - \epsilon^{-j\omega_0t})$$

By the use of Euler's identity the exponential terms may be replaced by their trigonometric equivalent, so that

$$i = \frac{\bar{E} - \frac{q_0}{C}}{L\omega_0} e^{-Rt/2L} \sin \omega_0 t \quad (5.14)$$

Equations (5.12) and (5.14) are different forms of the same equation. If $(1/LC) < (R^2/4L^2)$, ω_0 is imaginary by (5.13), and (5.12) is usually more convenient for calculating the current. Equation (5.12) may also be written in the hyperbolic form

$$i = \frac{\bar{E} - \frac{q_0}{C}}{L \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}} e^{-Rt/2L} \sinh \left(\sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}} t \right)$$

If $(1/LC) > (R^2/4L^2)$, ω_0 is real, and (5.14) is the more useful form.

A special case arises when $1/LC = R^2/4L^2$. Then ω_0 is zero, and (5.14) seems to be indeterminate,

$$i = \frac{\bar{E} - \frac{q_0}{C}}{L} e^{-Rt/2L} \frac{\sin \omega_0 t}{\omega_0}$$

However, by expanding $(\sin \omega_0 t)/\omega_0$ for small values of $\omega_0 t$,

$$\frac{\sin \omega_0 t}{\omega_0} = \frac{\omega_0 t - \frac{\omega_0^3 t^3}{3!} + \frac{\omega_0^5 t^5}{5!} - \dots}{\omega_0}$$

Thus the value of $(\sin \omega_0 t)/\omega_0$ approaches the value t as ω_0 approaches 0, and the current for the special case of $1/LC = R^2/4L^2$ becomes

$$i \left[\begin{array}{l} \text{critical} \\ \text{damping} \end{array} \right] = \frac{\bar{E} - \frac{q_0}{C}}{L} e^{-Rt/2L} t \quad (5.15)$$

The condition that $(1/LC) = (R^2/4L^2)$ is known as critical damping. Graphs of (5.12), (5.14), (5.15), are shown in Fig. 5.2 for constant values of L and C and three values of R .

If $R < 2\sqrt{L/C}$, the circuit is said to be underdamped; *i.e.*, the dissipation is small, and the current when the switch is closed is oscillatory. If $R > 2\sqrt{L/C}$, the circuit is overdamped and no oscillations occur. The condition $R = 2\sqrt{L/C}$ is called critical

damping. There is no reversal of current when the circuit is over-damped or critically damped.

Consider first the oscillatory transient represented by (5.14) and plotted in Fig. 5.3. The coefficient of the sine term establishes the envelope, or boundary, of the oscillations. The term $e^{-Rt/2L}$

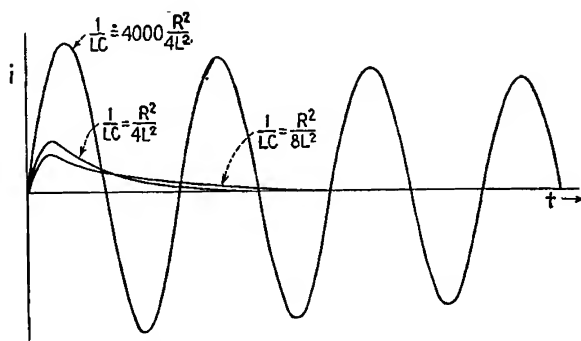


FIG. 5.2.—Graphs of equations (5.12), (5.14), (5.15) with $q_0 = 0$ plotted for three values of R ; L and C are held constant. All curves have the same initial slope E/L and enclose the same total area from zero to infinity of time. Areas above the time axis are considered positive, and areas below, negative.

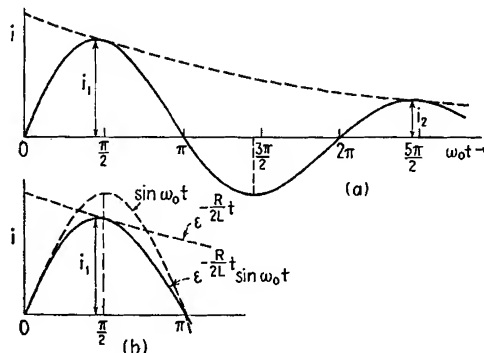


FIG. 5.3.—Graph of current versus time for an oscillatory transient showing (a) two successive maxima of current; (b) detail of the first maximum showing the difference between the positive peaks of the damped and undamped sinusoids.

is the damping term and accounts for the reduction in the amplitude of current peaks. If $2L/R$ is small compared with the period of oscillation, the oscillation boundaries come together rapidly and the oscillation dies out after relatively few cycles. If $2L/R$ is large compared with the duration of a cycle, the oscillation persists for many cycles. The ratio between the amplitudes of two successive

positive peaks i_1 and i_2 of the current curve in Fig. 5.3a is

$$\frac{i_1}{i_2} = \frac{\epsilon^{-Rt_1/2L} \sin \omega_0 t_1}{\epsilon^{-Rt_2/2L} \sin \omega_0 t_2}$$

where t_1 and t_2 are the times when the current has the maximum values shown. These times differ by one period,¹ so that

$$t_2 = t_1 + T_0$$

where

$$T_0 = \frac{2\pi}{\omega_0} = \frac{1}{f_0}$$

Then

$$\omega_0 t_2 = \omega_0 t_1 + 2\pi$$

and since

$$\begin{aligned} \sin \omega_0 t_1 &= \sin (\omega_0 t_1 + 2\pi) = \sin \omega_0 t_2 \\ \frac{i_1}{i_2} &= \frac{\epsilon^{-Rt_1/2L}}{\epsilon^{-\frac{R(t_1+T_0)}{2L}}} = \epsilon^{RT_0/2L} \end{aligned}$$

and

$$\log_e \frac{i_1}{i_2} = \frac{RT_0}{2L} = \delta \quad (5.16)$$

This value is the same for the ratio of any two successive positive (or negative) maxima. The logarithmic decrement δ is the ratio of the period of the oscillation to the time constant $2L/R$ of the envelope, Fig. 5.3a. Its reciprocal $1/\delta$ is approximately the number of complete oscillations of the current from the beginning of the transient to the time when the amplitude is reduced to 37 per cent of its initial amplitude (more correctly, to the time when $\epsilon^{-Rt/2L} = 0.37$). The logarithmic decrement $\delta = \pi R/L\omega_0 = \pi/Q$, where $Q = L\omega_0/R$ is the quality factor of the circuit. Thus Q/π is approximately the number of oscillations before the amplitude is reduced to 37 per cent of the maximum value.

Another interpretation of the quantity $1/\delta = Q/\pi$ can be expressed in terms of energy. Multiplying numerator and denominator of (5.16) by $\frac{1}{2}i_{\max}^2$ and inverting,

$$\frac{1}{\delta} = \frac{\frac{1}{2}L|i_{\max}|^2}{\frac{1}{2}R|i_{\max}|^2 \frac{T_0}{2}} \quad (5.17)$$

Now $\frac{1}{2}L|i_{\max}|^2$ is the maximum energy stored in the inductance during a cycle, and if the current is assumed to be sinusoidal (approxi-

¹ The instant of current maximum precedes the instant when $\sin \omega t = +1$ by a small amount, Fig. 5.3b, which increases with the damping. This amount is constant in each cycle and is equal to $CR/2$ when Q is large.

mately true if Q is large) $\frac{1}{2}|i_{\max}|^2 R = I_{\text{rms}}^2 R$, the average rate of energy dissipation in the circuit. Then $\frac{1}{2}i_{\max}^2 RT_0/2$ is the dissipation per half cycle of the oscillation. The fraction $1/\delta = 2L/RT_0$ is approximately the ratio of maximum stored energy in the inductance to the energy dissipated per half cycle. Since in a high- Q circuit practically all energy stored in the inductance at the instant of maximum current is transferred to the capacitance during a quarter cycle and back again to the inductance in the next quarter cycle, the maximum energy stored in the inductance is approximately the total stored energy in the circuit. Figure 5.6 shows the instantaneous plot of energies stored in the inductance and capacitance and their total as functions of time.

If the resistance of an LCR circuit is increased, the oscillation boundaries come closer together and at critical damping, $R = 2\sqrt{L/C}$, the transient ceases to be oscillatory. The critical value of R allows the most rapid charging of the capacitor from zero charge without "overshooting." Further increase of R results in the sort of current-time curve shown in Fig. 5.2. The initial slopes of all the current curves are the same if only R is varied, \bar{E} , L , and C remaining fixed. Furthermore, the areas under all the curves between zero and infinity are the same since the total charge moved around the circuit during the entire transient is independent of R .

Equation (5.12) can be reduced to a simpler form if

$$\frac{1}{LC} < \frac{R^2}{4L^2} \quad \text{or} \quad \frac{4L}{CR^2} < 1 \quad (5.18)$$

From (5.7),

$$\begin{aligned} k_1 &= -\frac{R}{2L} + \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}} \\ &= -\frac{R}{2L} + \frac{R}{2L} \sqrt{1 - \frac{4L}{R^2C}} \end{aligned}$$

Then, from (5.18),

$$k_1 = -\frac{R}{2L} \left(1 - 1 + \frac{2L}{R^2C} + \cdots \right) \doteq -\frac{1}{RC}$$

Similarly, from (5.8) and (5.18),

$$\begin{aligned} k_2 &= -\frac{R}{2L} - \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}} \\ &\doteq -\frac{R}{L} + \frac{1}{CR} \end{aligned}$$

Then (5.12) becomes

$$\begin{aligned}
 i &\doteq \frac{\bar{E} - \frac{q_0}{C}}{2L \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}} (\epsilon^{-\frac{t}{CR}} - \epsilon^{-\frac{R}{L}t + \frac{t}{CR}}) \\
 &\doteq \frac{\bar{E} - \frac{q_0}{C}}{\sqrt{R^2 - \frac{4L}{C}}} (\epsilon^{-\frac{t}{CR}} - \epsilon^{-\frac{R}{L}t + \frac{t}{CR}}) \quad (5.19)
 \end{aligned}$$

If L approaches zero, (5.19) reduces to

$$i = \frac{\bar{E} - \frac{q_0}{C}}{R} \epsilon^{-t/CR}$$

the equation for a CR circuit. Also, if C approaches infinity (no capacitor in the circuit), (5.19) becomes

$$i = \frac{\bar{E}}{R} (1 - \epsilon^{-Rt/L})$$

the equation for an LR circuit.

The charge on the capacitor in the LCR circuit is given by the solution of (5.2), viz.,

$$q = B_1 \epsilon^{k_1 t} + B_2 \epsilon^{k_2 t} + C\bar{E} \quad (5.20)$$

where k_1 and k_2 have the same values as in (5.7) and (5.8),

$$k_1 = -\frac{R}{2L} + \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}$$

and

$$k_2 = -\frac{R}{2L} - \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}$$

The constants B_1 and B_2 are evaluated from the boundary conditions. At $t = 0$, $q = q_0$, $i = 0$, so that $q_0 = B_1 + B_2 + C\bar{E}$. Differentiating (5.20) with respect to time,

$$i = B_1 k_1 \epsilon^{k_1 t} + B_2 k_2 \epsilon^{k_2 t} \quad (5.21)$$

By comparison with (5.12) it is evident that

$$B_1 = \frac{\bar{E} - \frac{q_0}{C}}{\left(2L \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}\right) k_1}$$

and

$$B_2 = \frac{-\left(\bar{E} - \frac{q_0}{C}\right)}{\left(2L\sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}\right)k_2}$$

and since the product k_1k_2 is

$$\left(-\frac{R}{2L} + \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}\right)\left(-\frac{R}{2L} - \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}\right) = \frac{1}{LC}$$

$$B_1 = \frac{\bar{E} - \frac{q_0}{C}k_2}{\left(2L\sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}\right)k_1k_2} = \frac{(q_0 - C\bar{E})\left(\frac{R}{2L} + \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}\right)}{2\sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}}}$$

which may be written

$$B_1 = \frac{q_0 - C\bar{E}}{2} \left(1 + \frac{1}{\sqrt{1 - \frac{4L}{R^2C}}}\right)$$

and similarly

$$B_2 = \frac{q_0 - C\bar{E}}{2} \left(1 - \frac{1}{\sqrt{1 - \frac{4L}{R^2C}}}\right)$$

Writing B_1 and B_2 in terms of $\omega_0 = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}$,

$$B_1 = \frac{q_0 - C\bar{E}}{2} \left(1 + \frac{R}{2L\omega_0}\right)$$

and

$$B_2 = \frac{q_0 - C\bar{E}}{2} \left(1 - \frac{R}{2L\omega_0}\right)$$

Writing out the entire equation for charge on the capacitor when the transient is oscillatory,

$$q = C\bar{E} + (q_0 - CE)e^{-Rt/2L} \cos \omega_0 t + (q_0 - C\bar{E}) \frac{R}{2L\omega_0} e^{-Rt/2L} \sin \omega_0 t \quad (5.22)$$

The graphs of charge vs. time for various values of R are shown in Fig. 5.4. Note that if $R/L\omega_0$ is small compared with unity

the curve of charge is approximately a damped cosine curve and thus lags the current curve, Fig. 5.3, by $\pi/2$ radians.

Since the expressions for current and charge have been developed in general form, the same equations hold for transients initiated by connecting a battery in the circuit with or without

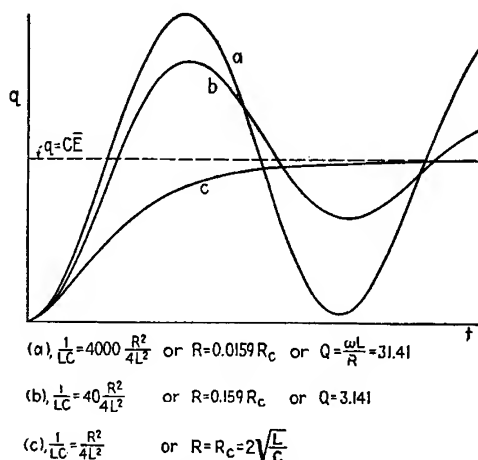


FIG. 5.4.—Graphs of equation (5.22) for various values of R .

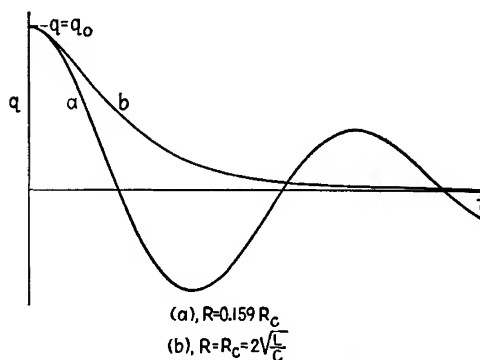


FIG. 5.5.—Curves of discharge of a capacitor through an LC circuit for two values of R .

an initial charge q_0 on the capacitor. Likewise, they hold where the transient is the discharge of an initially charged capacitor, with no battery in the circuit. The curves of charge and current for this case are determined by setting $\bar{E} = 0$ in (5.12) and (5.20). Two curves of q are shown in Fig. 5.5.

It is of interest to examine the energy stored in the inductance and in the capacitor during an oscillatory transient. The energy stored in the inductance is U_L , where

$$U_L = \frac{1}{2}Li^2$$

and in the capacitor U_C , where

$$U_C = \frac{1}{2} \frac{q^2}{C}$$

From (5.14) and (5.22), the total stored energy is

$$U = U_L + U_C = \frac{1}{2} \frac{q_0^2}{C} e^{-Rt/L} \left(1 + \frac{R}{2L\omega_0} \sin 2\omega_0 t + \frac{R^2}{2L^2\omega_0^2} \sin^2 \omega_0 t \right) \quad (5.23)$$

Graphs of the energy stored in the inductance and in the capacitance as well as of the total energy are shown in Fig. 5.6.

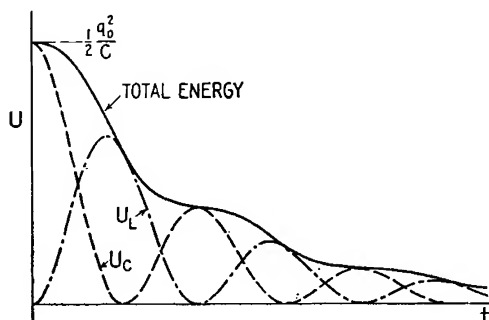


FIG. 5.6.—Graph of equation (5.23) showing: (solid line) total stored energy in the system during the discharge of a capacitor in an LCR circuit; (dash line) energy stored in the capacitor; (dot-dash line) energy stored in the inductor. The curves are plotted for the case where $\omega_0 L/R = \pi$.

In setting up a CR circuit it was assumed that such a circuit is physically possible. Actually, however, every circuit has some series inductance, and a more rigorous examination of the circuit is sometimes desirable. If a very small series inductance is present in a series CR circuit, the condition that $1 \gg 4L/R^2C$ is fulfilled and (5.19) is applicable. Graphs of the current in a series LCR circuit, Fig. 5.1, are shown in Fig. 5.7 for various values of L , including zero. If $L = 0$, the current jumps from zero to the value \bar{E}/R when the switch is closed, and the initial slope \bar{E}/L is infinite.

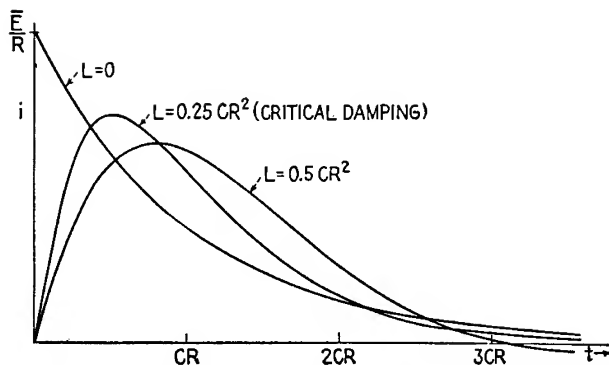


FIG. 5.7.—Showing the effect on the transient current of adding inductance to a series CR circuit.

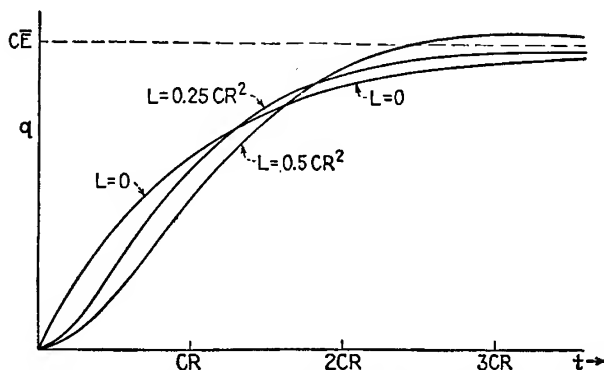


FIG. 5.8.—Showing the effect on the transient charge of adding inductance to a series CR circuit.

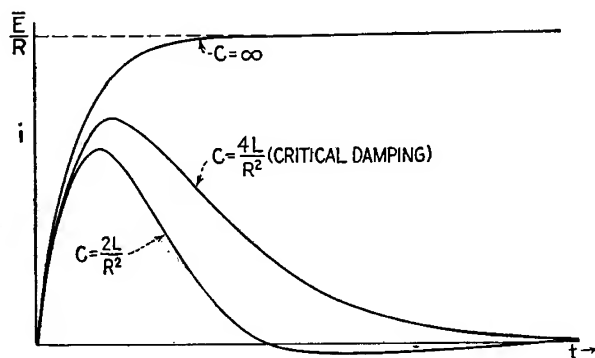


FIG. 5.9.—Effect on the current transient of varying the capacitance in a series LCR circuit, starting from infinite series capacitance or the equivalent of a series LR circuit.

The second curve for small L shows the form typical of actual CR circuits. This type of curve may be observed in any physical CR circuit where it is possible to resolve the transient by expanding the time scale.

Note that the curves for $L < R^2C/4$ ($L = R^2C/4$ is the critically damped case) all intersect the curve for $L = 0$ at two places. Another point of interest is the curve for $L = R^2C/2$. This curve intersects the zero axis in the shortest time and corresponds to the largest value of ω_0 , the natural frequency of the circuit, that can be obtained by variation of L . This particular value of L will appear in the discussion of amplifier compensation.¹ The corresponding curves of charge on the capacitor are shown in Fig. 5.8.

Families of curves of current vs. time for various values of series capacitance with L and R held constant are shown in Fig. 5.9.

6. Mechanical Analogues of the LCR Circuit.—The simplest analogue in mechanics to an LCR circuit is a mass suspended from a rigid support by a spring, Fig. 6.1. The force of gravity W acting downward corresponds to the battery voltage in the circuit of Fig. 5.1. The acceleration force ma acting in a direction opposite to the acceleration is analogous to the $L di/dt$ voltage. The friction force bv , where v is the velocity and b is a constant depending on the size and shape of the mass and on the viscosity of the medium in which the mass is suspended, is analogous to the voltage Ri and acts in a direction opposite to the velocity. The force ks

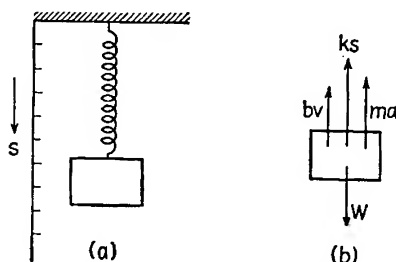


FIG. 6.1.—(a) Spring and weight system showing scale along which displacements s are measured; (b) free-body diagram showing the forces exerted on the mass as a dynamic system.

exerted by the spring, where k is the stiffness of the spring and s is the displacement, is the analogue of the voltage across the capacitor. Equating the sum of all these forces to zero,

$$W - ma - bv - ks = 0 \quad (6.1)$$

¹ See Chap. XIII. $L = R^2C/2$ corresponds to the condition where $N = 0.5$ for high-frequency compensation.

and since the acceleration $a = d^2s/dt^2$ and the velocity $v = ds/dt$, (6.1) may be written

$$W - m \frac{d^2s}{dt^2} - b \frac{ds}{dt} - ks = 0 \quad (6.2)$$

This equation is of the same form as (5.2) and has the same form of solution. If the mass is disturbed from its equilibrium position, given by the condition $W = ks$, it will oscillate up and down about the equilibrium position if the experiment is performed in air where b is small. If the experiment is repeated in a medium of sufficient viscosity, the motion becomes critically damped or overdamped. The equation of the angular displacement of a pendulum is similar to (6.2) when the displacement is small.

Another excellent analogue of the oscillatory case is the oscillation of a balance-wheel and hairspring combination used to regulate the rate of unwinding of the mainspring of a clock or watch. When used in a clock, the pendulum or the balance-wheel system is an example of an oscillatory system in which the damping is very small, *i.e.*, the decrement is small or the Q is high. An example of critical damping is met in the pointer on an indicating ammeter or voltmeter. It is undesirable to have the pointer oscillate about the final indication when a given current is sent through the instrument. On the other hand, the damping may be too high, causing the pointer to be sluggish in reaching its final reading. The critical damping gives the most rapid arrival at the final reading without overshooting. Actually, the damping is adjusted usually to be slightly less than critical so that the pointer overshoots slightly but does not make more than one visible oscillation. A spring balance is not damped ordinarily and has to be stopped from oscillating by hand in order to get a reading quickly.

Vibrations of plucked strings on musical instruments, vibrations of drumheads, etc., are further examples of oscillatory transients in mechanical systems, although the conditions are not analogous to the simple *LCR* circuit.

The *LCR* combination having a high Q is used to fix the frequency of oscillation of an oscillator just as the extremely high Q systems of the pendulum or balance wheel are used to govern the frequency of a clock or watch. The oscillation of a piezoelectric crystal is another example of an extremely high- Q electro-mechanical system whose oscillations can be maintained electrically and used in very accurate timing devices.

7. Square-wave Testing.—Another application of transient equations is in the analysis of the performance of an amplifier. With an input voltage of square waveform together with an examination of the output waveform it is possible to determine the phase and frequency distortion of the amplifier.

Then, if $t = 0$ at the beginning of any positive or negative alternation, (3.5) applies and

$$q \Big|_{t=T/2} = EC(1 - e^{-T/2CR}) + q_0 e^{-T/2CR} \quad (7.1)$$

Since the charge at the end of an alternation is the negative of that at the beginning (on the assumption that many cycles have elapsed),

$$q \Big|_{t=T/2} = -q_0$$

and (7.1) yields

$$\begin{aligned} q \Big|_{t=T/2} &= EC \frac{1 - e^{-T/2CR}}{1 + e^{-T/2CR}} = -q_0 \\ &= EC \tanh \frac{T}{4CR} \end{aligned} \quad (7.2)$$

When a voltage of square waveform is applied to a simple CR circuit, the steady state is a series of recurring transients, Fig. 4.1. Consider the graph of Fig. 4.1c. Since in the steady state the flow of charge in one direction equals the flow of charge in the opposite direction (there being no steady-state direct current resulting from a square-wave voltage), the maximum voltage across the capacitor in the positive direction is equal to the maximum voltage across the capacitor in the negative direction. Hence the charge on the capacitor at the end of one of the recurring transients is equal and opposite to the charge at the beginning. This charge is the charge left on the capacitor at the end of each half cycle of operation and is positive at the end of a positive alternation of the square wave. Substitution of this value as initial charge in (3.4) gives the analytic form of the current. At the beginning of each positive half cycle, the initial charge is given by (7.1) with a negative sign. The current then is

$$\begin{aligned} i &= \frac{E}{R} e^{-t/CR} + \frac{E}{R} \left(\frac{1 - e^{-T/2CR}}{1 + e^{-T/2CR}} \right) e^{-t/CR} \\ &= \frac{2E}{R(1 + e^{-T/2CR})} e^{-t/CR} \end{aligned} \quad (7.3)$$

Figure 4.1 shows the applied square wave and the resultant voltages across the resistor and the capacitor for several assumed time constants. Note that the initial value of the voltage across the resistor is approximately twice the applied square-wave voltage for $CR \ll T/2$.

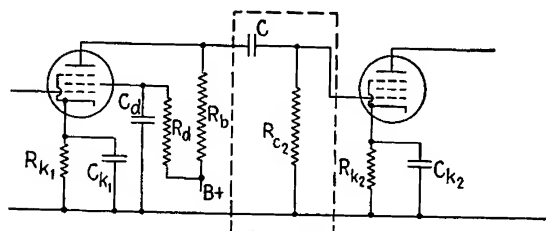


FIG. 7.1.—Typical RC -coupled amplifier; the coupling capacitor and grid resistor are enclosed by the dashed line.

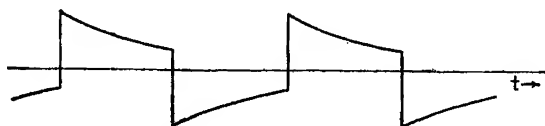


FIG. 7.2.—Typical distortion of a square wave caused by distortion of the low-frequency components.

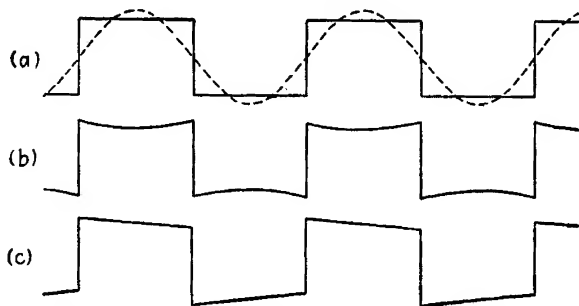


FIG. 7.3.—(a) Square wave and its fundamental component; (b) effect of only reducing the amplitude of the fundamental component; (c) effect of only shifting the phase of the fundamental component.

Low-frequency Distortion.—Consider first the coupling capacitor and grid resistor of an amplifier, Fig. 7.1. Suppose a square voltage wave is developed across the load resistor R_b ; then the voltage at the grid of the next tube might have the form shown in Fig. 7.2. The output-voltage waveform is a series of recurring transients. The equation of each curved part is given by (7.3) multiplied by R_c . This is an obvious distortion of the square-wave voltage due to the fact that the low-frequency components are reduced in

amplitude and shifted in phase. Figure 7.3*a* shows the square-wave voltage and the fundamental component of voltage. Suppose that the fundamental component is reduced in amplitude but not shifted in phase. Then the resultant waveform would have the appearance of Fig. 7.3*b*. If, on the other hand, the fundamental component is shifted slightly in phase, the effect is to raise one corner of the square-wave voltage and depress the other corner, Fig. 7.3*c*. This is the sort of distortion encountered in RC -coupled amplifiers.

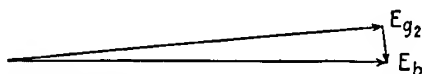


FIG. 7.4.—Voltage vector diagram for the fundamental component of the square wave; $\tan \theta = 0.1$; $\theta = 5^\circ 40'$.

There is a rule of thumb, often used in amplifier design, which states that the lower half-power frequency f' should be one-tenth the lowest-frequency component of the signal. The lower half-power angular frequency in an RC -coupled amplifier is $\omega' \doteq 1/(R_c C)$. The tangent of the angle of lead for the voltage developed across R_{c2} is $\tan \theta = X_c/R_{c2} = 1/\omega C R_{c2}$; at the half-power frequency $\omega = \omega'$, $\tan \theta = 1$. Suppose that ω_1 , the fundamental angular frequency, is ten times cutoff frequency, that is, that $\omega_1 = 10\omega'$. This means that R_{c2} and C have been so chosen that ω' is one-tenth the fundamental frequency of the applied square wave ω_1 . Under

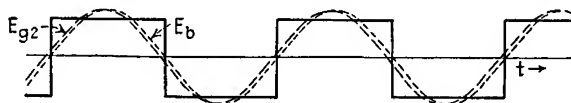


FIG. 7.5.—Angular shift of the fundamental component corresponding to the vector diagram of Fig. 7.4 and causing the distortion shown in Fig. 7.3*c*.

these conditions the value of $\tan \theta_1$ for the fundamental component will be $\frac{1}{10}$. The angle whose tangent is $\frac{1}{10}$ is $5^\circ 40'$ so that $\cos \theta = 0.995$ and $\sin \theta = 0.099$. As shown in Fig. 7.4, E_{g2} , the voltage applied to the grid of the succeeding stage, leads E_b , the fundamental voltage developed across the resistor R_b , by $5^\circ 40'$. Since $E_{g2} = E_b \cos \theta = 0.995E_b$, the effect of the coupling network in reducing the magnitude of the fundamental voltage is negligible under the assumed conditions, being of the order of one-half of 1 per cent.

However, the $5^\circ 40'$ phase shift has the effect shown in Figs. 7.5 and 7.3*c* of raising the leading edge of the square wave by an

amount proportional to the sine of $5^\circ 40'$. Since the amplitude of the fundamental component is $4/\pi$ times the square-wave amplitude, the leading edge is lifted by a factor $4/\pi \sin 5^\circ 40' = 0.126$. Thus the leading edge is raised by 12.6 per cent of its undistorted value, and the trailing edge is lowered by the same amount, causing the output wave to have the appearance shown in Fig. 7.3c. This is typical of low-frequency distortion in an RC -coupled amplifier, and it is evident that the first visible effects of low-frequency distortion are largely due to phase shift and not to reduction in amplitude.

High-frequency Distortion.—The equivalent high-frequency circuit of an amplifier can be reduced to the form of Fig. 7.6. The input is a square-wave current of amplitude g_mE_{g1} , where E_{g1} is the square-wave voltage applied to the input grid and g_m is the transconductance of the tube. In the diagram, C_s is the total shunting capacitance and R_{par} is the equivalent resistance of the parallel combination of plate resistance r_p , load resistance R_b , and the grid-leak resistance R_{c2} of the following stage. Let i_R be the variational current through the equivalent resistance R_{par} and i_C the current through the capacitance C_s . The output voltage developed across R_{par} is equal to the voltage across C_s , so that

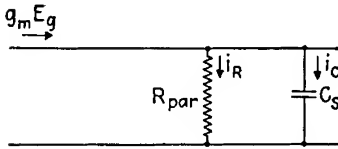


FIG. 7.6.—High-frequency equivalent circuit of an amplifier.

$$E_{g2} = i_R R_{par} = \frac{q}{C_s} \quad (7.4)$$

Differentiating (7.4),

$$R_{par} \frac{di_R}{dt} = \frac{1}{C_s} \frac{dq}{dt} = \frac{1}{C_s} i_C \quad (7.5)$$

Now

$$i_C = g_m E_g - i_R$$

Substituting this value in (7.5) and separating the variables,

$$\frac{di_R}{g_m E_g - i_R} = \frac{1}{C_s R_{par}} dt \quad (7.6)$$

Since $g_m E_g$ is constant for the duration of a square-wave alternation, it may be considered as a constant in the integration over the duration of the half cycle.

$$\int_{(i_R)_{t=0}}^{i_R} \frac{di_R}{g_m E_g - i_R} = \frac{1}{C_s R_{par}} \int_0^t dt$$

whence

$$\log_e \frac{i_p - i_R}{i_p - (i_R)_{t=0}} = - \frac{1}{C_s R_{par}} t \quad (7.7)$$

or, expressed in exponential form,

$$i_R = i_p(1 - e^{-t/C_s R_{par}}) + (i_R)_{t=0} e^{-t/C_s R_{par}} \quad (7.8)$$

If a number of cycles have taken place, the condition that the current at the end of each positive half cycle be equal to the negative of the current at the end of a negative half cycle is fulfilled and the current at the end of a positive half cycle will be given by substituting $T/2$ for the time, where T is the period of the square wave. Then

$$(i_R)_{t=0} = i_p(1 - e^{-T/2R_{par}C_s}) - (i_R)_{t=0} e^{-T/2R_{par}C_s}$$

or

$$(i_R)_{t=0} = i_p \frac{(1 - e^{-T/2R_{par}C_s})}{(1 + e^{-T/2R_{par}C_s})} = i_p \tanh \frac{T}{4R_{par}C_s}$$

and the current through R_{par} at any time t of the cycle will be given by (7.9).

$$i_R = i_p(1 + e^{-t/R_{par}C_s}) - i_p \frac{(1 - e^{-T/2R_{par}C_s})}{(1 + e^{-T/2R_{par}C_s})} e^{-t/R_{par}C_s}$$

which reduces to

$$i_R = i_p \left(1 - \frac{2e^{-t/R_{par}C_s}}{1 + e^{-T/2R_{par}C_s}} \right) \quad (7.9)$$

and if $T/2$, the duration of a half cycle of the square wave, is much greater than the time constant of the circuit, $T/2 \gg R_{par}C_s$, then $e^{-T/2R_{par}C_s}$ is a very small quantity compared with unity and $i_R \doteq i_p(1 - 2e^{-t/R_{par}C_s})$.

The voltage across the load, which is the voltage applied to the next stage of the amplifier, is

$$e_{g2} = i_R R_{par}$$

or

$$e_2 = E_{g1} g_m R_{par} \left(1 - \frac{2e^{-t/R_{par}C_s}}{1 + e^{-T/2R_{par}C_s}} \right) \quad (7.10)$$

Figure 7.7 shows the response according to (7.10) of the circuit of Fig. 7.6 when E_{g1} has a square-wave form. For Fig. 7.7a, the

circuit parameters R and C_s are adjusted so that the half-power frequency ($\omega'' = 1/R_{par}C_s$) equals the frequency of the square wave. For Fig. 7.7b, the time constant $R_{par}C_s$ has been reduced so that the half-power frequency ω'' is 10 times the frequency of the square wave. The output-wave form now resembles the input square wave. When $R_{par}C_s$ is further reduced so that the upper half-power frequency is 100 times the frequency of the square wave, the reproduction is excellent, Fig. 7.7c.

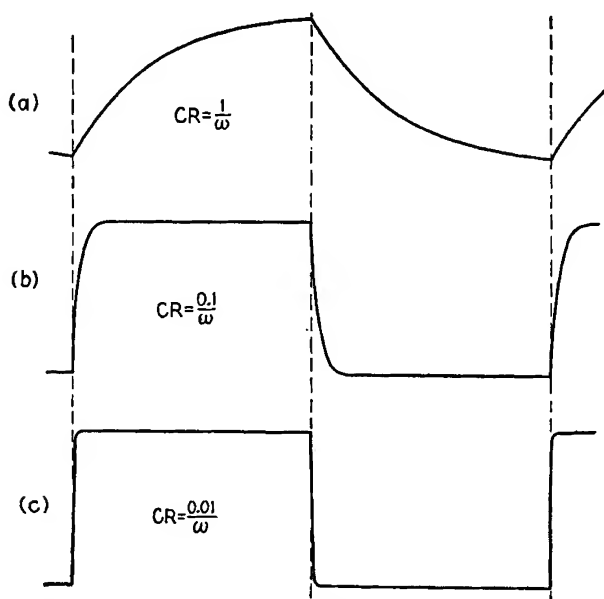


FIG. 7.7.—Distortion of a square wave caused by distortion of the high-frequency components.

The effects of high-frequency distortion of a square wave may be presented in another way: The leading edge of the square wave will have a certain shape dictated by the time constant $R_{par}C_s$ of the circuit. If the half period of the square wave is short compared with the time constant of the circuit, the output wave will be distorted; but if the half period is long compared with the time constant, then the output wave will be reasonably "square."

Figure 7.8 shows the equivalent circuit of a video-frequency amplifier, series-compensated by the use of inductance L_b in series with the plate-load resistor R_b . The transient analysis of this circuit is somewhat involved but follows the general method used

for any LCR circuit. The boundary conditions for a square-wave input are that the current through R_b and the charge on the capacitance C_s at the end of a positive half cycle are the negatives of the respective values at the end of a negative half cycle. The resulting

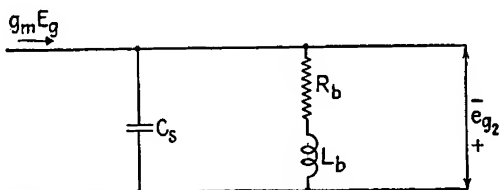


FIG. 7.8.—Equivalent circuit of a series-compensated video amplifier stage.

output voltage for a positive half cycle, taken after many cycles have elapsed, is

$$e_{g2} = g_m E_{g1} R_b (1 + F_1 e^{k_1 t} + F_2 e^{k_2 t}) \quad (7.11)$$

where k_1 and k_2 are as defined previously and

$$F_1 = \frac{1 + C_s R_b k_2}{C_s R_b \sqrt{\frac{R^2}{4L^2} \frac{1}{LC}} (1 + e^{k_1 \tau/2})}$$

and

$$F_2 = \frac{1 + C_s R_b k_1}{C_s R_b \sqrt{\frac{R^2}{4L^2} \frac{1}{LC}} (1 + e^{k_2 \tau/2})}$$

The form of the negative half cycle is given by the negative of (7.11). Figure 7.9 shows the effect of the compensating inductance on the leading edge of the output wave. The initial slope of the leading edge of the output wave is not dependent on L_b . The curves are plotted for $L_b = 0$, $L_b = 0.25CR^2$, and $L_b = 0.5CR^2$.^{*} Since the usual values of L_b lie between $0.3CR^2$ and $0.5CR^2$, the circuit is oscillatory. $L_b = 0.25CR^2$ corresponds to critical damping.

Another application of square-wave testing arises in connection with the adjustment of a compensated voltage divider. Figure 7.10a shows such a circuit. Because of the input capacitance of the vacuum tube it is necessary to provide the voltage divider $R_1 R_2$ with a compensating capacitance C_c to make the voltage division independent of frequency. The required C_c is given by $R_1 C_c = R_2 C_v$. The proper adjustment of C_c is easily recognized

^{*} These values correspond to N equal to 0, 0.25, 0.5, Chap. XIII.

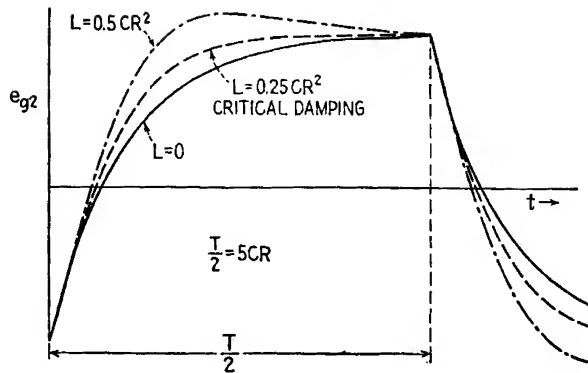


FIG. 7.9.—Graph of equation (7.11) for a positive half cycle and part of the next negative half cycle of the input square wave.

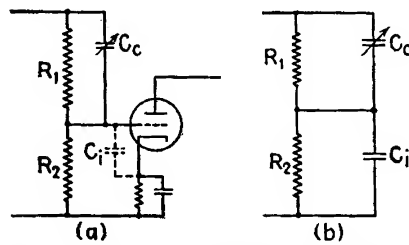


FIG. 7.10.—(a) Input circuit with compensating capacitance C_c ; (b) equivalent simplified circuit.

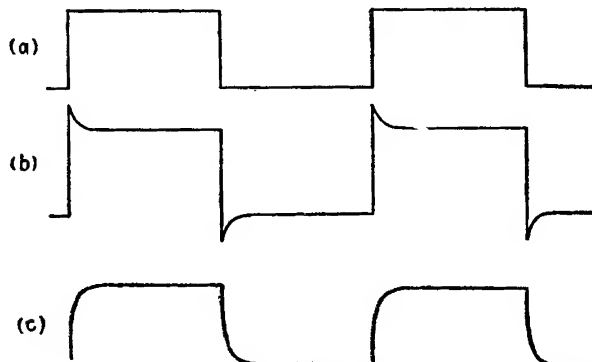


FIG. 7.11.—Voltage across R_2 of Fig. 7.10 for (a) proper adjustment of C_c , (b) C_c too large, (c) C_c too small.

when a square-wave voltage is applied to the divider and the output waveform is examined on the oscillograph. Figure 7.11*a* shows the waveform on the grid for correct adjustment of C_c . The waveform for too large a value of C_c is shown in Fig. 7.11*b* and for too small a value of C_c in Fig. 7.11*c*. The period of the square wave used for adjustment of C_c should be of the order of magnitude of the time constant $C_i R_1 R_2 / (R_1 + R_2)$ of the uncompensated divider.

In all applications of square waves for test purposes it must be borne in mind that the internal impedance of the square-wave generator must be taken into account in any calculation of output waveforms.

By calibration of C_c in Fig. 7.10*b* and using known resistors R_1 and R_2 , the input capacitance of a given vacuum tube can be measured using the square-wave response as an indication of balance of the equation $R_2 C_i = R_1 C_c$.

CHAPTER VII

COUPLED CIRCUITS

1. Introduction.—Two or more complete circuits which are connected as parts of one network so that they react upon one another are called *coupled circuits*. When the circuits have both inductive and capacitive reactances, the interaction between the circuits imparts to them interesting and desirable properties that are quite different from the properties of a single tuned circuit.

These special and useful properties of coupled circuits make them applicable to many types of electronic equipment. For example, two coupled circuits are used often as coupling networks

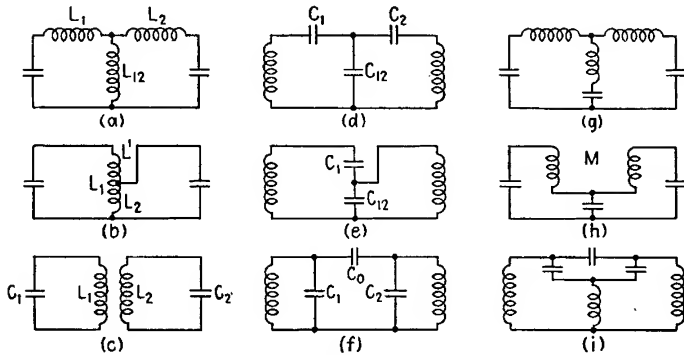


FIG. 1.1.1.—Various types of coupled circuits.

in receivers because of the wide transmission band that the coupled system possesses. As another example, coupled circuits are used generally as coupling systems from the power-amplifier stage of transmitters to the antenna, in which case they are used to match impedances and to suppress harmonics. Because of the many applications of coupled systems every radio and electronics engineer should be thoroughly familiar with their characteristics and properties.

In the treatment that follows, both the transient response and the response to forced oscillations will be discussed. A familiarity

with the transient response is as essential as in connection with single circuits, for any sudden change in the existing condition sets up transient oscillations.

Two circuits may be magnetically coupled as in Figs. 1.1*a*, *b*, *c* or capacitively coupled as in Figs. 1.1*d*, *e*, *f*. They may also be coupled by a resistance common to the two circuits. Besides these three pure types of coupling, combinations of the three types may be used, as in Figs. 1.1*g*, *h*, *i*. A network of two coupled circuits can be transformed into the general representation of Fig. 1.2.

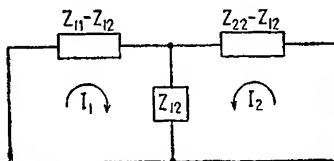


FIG. 1.2.—General representation of two coupled circuits.

2. Coefficient of Coupling.—The degree of coupling is expressed by a *coefficient of coupling* denoted by k . There must be a separate k for that portion of the coupling involving each type of circuit element. The value of k for the circuits shown in Fig. 1.1, expressed in terms of the general circuit of Fig. 1.2, is

$$k = \frac{|X_{12}|}{\sqrt{X_{11}X_{22}}} \quad \text{for one type of circuit element} \quad (2.1)$$

For resistance coupling the reactances in (2.1) would be replaced by resistances. The magnitude of k is independent of frequency provided that the circuit elements are constant; and k varies in magnitude from zero to unity. The values of k for the circuits shown in Fig. 1.1 are

$$k = \frac{L_{12}}{\sqrt{(L_1 + L_{12})(L_2 + L_{12})}} \quad \text{Fig. 1.1a} \quad (2.2)$$

$$= \frac{L_1 + L_2 - L'}{2\sqrt{L_1L_2}} \quad \text{Fig. 1.1b} \quad (2.3)$$

$$= \frac{M}{\sqrt{L_1L_2}} \quad \text{Fig. 1.1c} \quad (2.4)$$

$$= \sqrt{\frac{C_1C_2}{(C_1 + C_{12})(C_2 + C_{12})}} \quad \text{Fig. 1.1d} \quad (2.5)$$

$$= \sqrt{\frac{C_1}{C_1 + C_{12}}} \quad \text{Fig. 1.1e} \quad (2.6)$$

$$= \frac{C_0}{\sqrt{(C_1 + C_0)(C_2 + C_0)}} \quad \text{Fig. 1.1f} \quad (2.7)$$

I. TRANSIENT OSCILLATIONS

3. General Considerations.—When two coupled circuits are disturbed by any sudden electrical change, such as discharging one of the capacitors through its circuit, transient oscillations are set up. These oscillations plotted against time are more complex than the simple damped train of oscillations of a single circuit of Fig. 5.2, Chap. VI. Typical time plots of the oscillations of two coupled circuits are shown in Fig. 3.1. Curve 1 gives the voltage across the primary capacitor; curve 2 gives the voltage across the secondary capacitor.

The amplitude of the voltage in each circuit rises and falls, forming groups of oscillations called *beats*. The energy $C_1 E_1^2/2$,

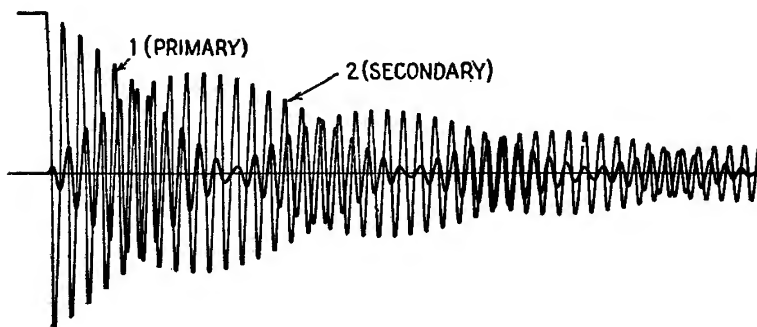


FIG. 3.1.—Oscillations of two coupled circuits.

originally present in C_1 , is transferred partly or wholly to the secondary circuit and then back to the primary circuit, and so on, back and forth, as the energy is gradually dissipated in the resistances of the circuits.

The waveforms of Fig. 3.1 are precisely the same as the form of the resultant sound wave when two tuning forks of slightly different pitch vibrate simultaneously. The beats of Fig. 3.1 therefore indicate the presence of *two* frequencies in each of the circuits. The closer the coupling, the fewer the number of oscillations in each beat and the greater the difference between the two frequencies of oscillation. If m denotes the number of oscillations in one beat and if each of the circuits individually has the same natural frequency, then

$$k \doteq \frac{1}{m} \quad (3.1)$$

4. Frequencies of Transient Oscillations.—The magnitudes of the frequencies of free oscillations of two coupled circuits cannot easily be derived for circuits having appreciable resistance. However, since in a single circuit the resistance, if small, makes little difference in the frequency of free oscillation, (5.13), Chap. VI, it would be logical to suppose that the frequencies of free oscillation of coupled circuits would be affected only slightly by small resistances in the circuits, and this is the fact. Therefore, the frequencies will be stated and derived for resistanceless circuits.

The derivation is given in Sec. 5; the formulas are given here. Let ω^* and ω^{**} be the angular frequencies of the low and the high frequency of free oscillations. Then, for two *magnetically* coupled circuits,

$$\left. \begin{matrix} \omega^* \\ \omega^{**} \end{matrix} \right\} = \sqrt{\frac{\omega_1^2 + \omega_2^2 \pm \sqrt{(\omega_1^2 + \omega_2^2)^2 - 4\omega_1^2\omega_2^2(1 - k^2)}}{2(1 - k^2)}} \quad (4.1)$$

In this expression

$$\omega_1 = \frac{1}{\sqrt{L_1 C_1}} \quad \text{and} \quad \omega_2 = \frac{1}{\sqrt{L_2 C_2}}$$

ω_1 and ω_2 being the angular frequencies of free oscillation of the circuits when alone and unaffected by each other.

It is often better to deal with the *wavelength* λ in free space, which corresponds to an angular velocity ω . Since the velocity of an electromagnetic wave in free space is the velocity of light, denoted by $v_c = 3 \cdot 10^8$ meters/sec, then

$$\lambda = \frac{v_c}{f} = \frac{2\pi v_c}{\omega} \quad (4.2)$$

Equation (4.1) converted to wavelengths becomes

$$\left. \begin{matrix} \lambda^* \\ \lambda^{**} \end{matrix} \right\} = \sqrt{\frac{\lambda_1^2 + \lambda_2^2 \pm \sqrt{(\lambda_1^2 + \lambda_2^2)^2 - 4\lambda_1^2\lambda_2^2(1 - k^2)}}{2}} \quad (4.3)$$

for two *magnetically* coupled circuits.

Equations (4.1) and (4.3) are limited to magnetically coupled circuits. When the circuits are capacitively coupled, the wavelengths are those obtained by replacing the symbol ω by λ in (4.1) and the angular frequencies are those obtained by replacing the symbol λ by ω in (4.3), the subscripts remaining unchanged and the superscripts * and ** being interchanged in both cases.

A better conception of the meaning of (4.3) can be obtained from its graphical representation in Fig. 4.1, where the coordinates are given in ratios of the wavelengths to the natural primary-circuit wavelength λ_1 . The 45-degree line and the horizontal line, which intersect at the point above $\lambda_2/\lambda_1 = 1$ on the horizontal axis,

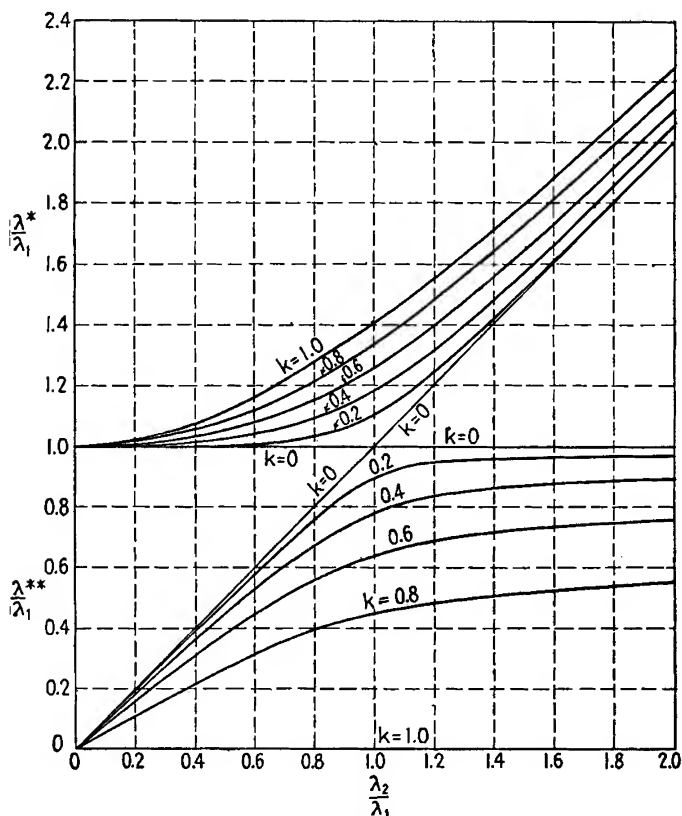


FIG. 4.1.—Wavelengths of a freely oscillating coupled system.

represent the free-oscillation wavelengths, or natural wavelengths, for zero coupling.

The curves of Fig. 4.1 show that as the coefficient of coupling of two circuits is increased the wavelengths of the free oscillations move apart from the values which would exist in the circuits for zero coupling. When the coefficient of coupling approaches unity, one of the free-oscillation wavelengths approaches zero, or its frequency approaches infinity. However, the energy associated

with this frequency approaches zero, so that, in the limit, currents of this frequency do not appear.

When the two circuits have the same natural wavelength, $\lambda_2/\lambda_1 = 1$, and (4.3) reduces to

$$\frac{\lambda^*}{\lambda_1} \text{ and } \frac{\lambda^{**}}{\lambda_1} = \sqrt{1 \pm k} \quad (4.4)$$

This is the case of greatest practical interest since circuits in receivers and transmitters are tuned usually for $\lambda_2 = \lambda_1$. The oscillations of Fig. 3.1 are for this condition.

5. Derivation of Frequencies of Free Oscillation of Magnetically Coupled Circuits.—The differential equations expressing the voltages in the circuits of Fig. 1.1c, resistances being neglected, are¹

$$L_1 \frac{di_1}{dt} + M \frac{di_2}{dt} + \frac{q_1}{C_1} = 0 \quad (5.1)$$

$$L_2 \frac{di_2}{dt} + M \frac{di_1}{dt} + \frac{q_2}{C_2} = 0 \quad (5.2)$$

Differentiating (5.1) and (5.2) and substituting $dq_1/dt = i_1$ and $dq_2/dt = i_2$,

$$L_1 \frac{d^2 i_1}{dt^2} + M \frac{d^2 i_2}{dt^2} + \frac{i_1}{C_1} = 0 \quad (5.3)$$

$$L_2 \frac{d^2 i_2}{dt^2} + M \frac{d^2 i_1}{dt^2} + \frac{i_2}{C_2} = 0 \quad (5.4)$$

Since there are no resistances in the circuits, the currents will be sinusoidal in form with no damping. Hence, assume that

$$i_1 = \hat{I}_1 \sin \omega t \quad (5.5)$$

$$i_2 = \hat{I}_2 \sin \omega t \quad (5.6)$$

where the ω 's are unknown and will be evaluated. Substituting these expressions in (5.3) and (5.4),

$$\left(\omega^2 L_1 - \frac{1}{C_1} \right) \hat{I}_1 \sin \omega t = -\omega^2 M \hat{I}_2 \sin \omega t \quad (5.7)$$

$$\left(\omega^2 L_2 - \frac{1}{C_2} \right) \hat{I}_2 \sin \omega t = -\omega^2 M \hat{I}_1 \sin \omega t \quad (5.8)$$

¹ Note that the algebraic sign before M may be either plus or minus, but the sign chosen determines whether magnetic flux caused by positive flow of secondary current is in the same direction as or the opposite direction to the flux caused by positive flow of primary current. The plus sign adopted fixes the convention that positive currents produce aiding magnetic fields.

Multiplying (5.7) by (5.8) and canceling $\hat{I}_1 \hat{I}_2 \sin^2 \omega t$,

$$\left(\omega^2 - \frac{1}{L_1 C_1}\right) \left(\omega^2 - \frac{1}{L_2 C_2}\right) = \frac{\omega^4 M^2}{L_1 L_2} \quad (5.9)$$

or

$$(\omega^2 - \omega_1^2)(\omega^2 - \omega_2^2) = \omega^4 k^2 \quad (5.10)$$

Solving (5.10) for ω gives (4.1).

Dividing (5.7) by (5.8) gives the ratio \hat{I}_2/\hat{I}_1 , or the relative magnitudes of the currents of the same frequency in the two circuits.¹

II. FORCED OSCILLATIONS

6. General Considerations.—By “forced oscillations” is meant the steady-state conditions of current, voltage, etc., at any point

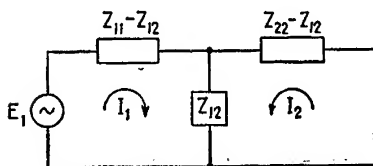


FIG. 6.1.—General form of two coupled circuits.

resulting from the application of a sustained, periodically varying emf at a specified point in the network. The periodically varying emf may have any waveform; but since any periodic waveform may be resolved into a series of sinusoidal components,

the usual procedure in studying forced oscillations is to find the response of the network to any single sinusoidal emf.

The combination of any two coupled circuits may be represented by the more general network of Fig. 1.2. If a sinusoidal emf E_1 of angular frequency ω is impressed in series with the first circuit as in Fig. 6.1, then (2.32) and (2.33) of Chap. V express the electrical conditions in the circuits.

$$I_1 Z_{11} + I_2 Z_{12} = E_1 \quad (6.1)$$

$$I_1 Z_{12} + I_2 Z_{22} = 0 \quad (6.2)$$

Solving (6.1) and (6.2) for I_1 and I_2 ,

$$I_1 = \frac{E_1}{Z_{11} - \frac{Z_{12}^2}{Z_{22}}} \quad (6.3)$$

$$I_2 = - \frac{Z_{12} E_1}{Z_{11} Z_{22} - Z_{12}^2} \quad (6.4)$$

¹ E. L. CHAFFEE, Amplitude Relations in Coupled Circuits, *Proc. I.R.E.*, 4, 283, 1916.

Equations (6.3), (6.4) are general and apply to any of the circuits in Fig. 1.1 when the proper values for Z_{11} , Z_{12} , Z_{22} are substituted in the equations.

For two magnetically coupled circuits, Fig. 6.2,

$$Z_{11} = R_1 + j\left(\omega L_1 - \frac{1}{\omega C_1}\right) = R_1 + jX_1 \quad (6.5)$$

$$Z_{22} = R_2 + j\left(\omega L_2 - \frac{1}{\omega C_2}\right) = R_2 + jX_2 \quad (6.6)$$

$$Z_{12} = j\omega M \quad (6.7)$$

Impedance Z_{12} , which is numerically equal to the voltage introduced into the secondary circuit by unit current in the primary

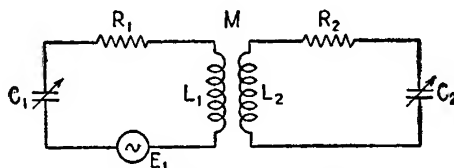


FIG. 6.2.—Two magnetically coupled circuits.

circuit, is $j\omega M$ for this case. Then (6.3) and (6.4) become

$$I_1 = \frac{E_1}{Z_{11} + \frac{\omega^2 M^2}{Z_{22}}} \quad (6.8)$$

$$I_2 = \frac{-j\omega M E_1}{Z_{11} Z_{22} + \omega^2 M^2} \quad (6.9)$$

7. Equivalent Impedance of Coupled Circuits.—The equivalent impedance of the network as presented to the generator E_1 for the general case of Fig. 6.1 is given by the denominator of (6.3). Considering, however, only the special case of Fig. 6.2, the equivalent primary circuit impedance Z_{11}' is given by the denominator of (6.8), or

$$Z_{11}' = Z_{11} + \frac{\omega^2 M^2}{Z_{22}} \quad (7.1)$$

This equivalent impedance consists of two terms; the first term Z_{11} is the impedance of the primary circuit alone; the second term $\omega^2 M^2 / Z_{22}$ is the impedance “reflected” into the primary circuit from the secondary circuit.

Equation 7.1 can be expanded so as to give the resistive and reactive components of Z_{11}' , thus:

$$Z_{11}' = \left(R_1 + \frac{\omega^2 M^2 R_2}{R_2^2 + X_2^2} \right) + j \left(X_1 - \frac{\omega^2 M^2 X_2}{R_2^2 + X_2^2} \right) \quad (7.2)$$

Equation (7.2) shows that the reactance X_2 has the effect of a reactance of the opposite sign added in series in the primary circuit. For example, a metallic shield placed over or near a coil acts as a short-circuited secondary circuit having a positive reactance. Its effect upon the coil is to *increase* its equivalent resistance and to *decrease* its equivalent inductive reactance.

If the secondary circuit is tuned to resonance, X_2 is zero, and the only effect of the secondary circuit is to add a resistance of $\omega^2 M^2 / R_2$ ohms in the primary circuit. This is an important case.

8. Magnitude of Secondary Current.—Generally the greatest interest centers in the magnitude of the secondary current and its dependence upon the coupling and the tuning of the two circuits. In order to study the secondary current the expression for its magnitude must be obtained. This expression can be derived from (6.9) by applying the usual method of finding the magnitude of a complex quantity. Thus

$$|I_2| = \frac{|\omega M E_1|}{\sqrt{(R_1 R_2 - X_1 X_2 + \omega^2 M^2)^2 + (R_1 X_2 + R_2 X_1)^2}} \quad (8.1)$$

where, from (6.5) and (6.6),

$$X_1 = \omega L_1 - \frac{1}{\omega C_1} \quad (8.2)$$

$$X_2 = \omega L_2 - \frac{1}{\omega C_2} \quad (8.3)$$

Generally the adjustable factors in a coupled system, such as that of Fig. 6.2, are M , C_1 , C_2 . Changing C_1 or C_2 varies X_1 or X_2 as shown by (8.2) and (8.3). The magnitude of I_2 can then be expressed either in terms of M , X_1 , X_2 or in terms of M , C_1 , C_2 . The state of tuning of the circuits can also be expressed in terms of the series-resonant frequencies of the primary and secondary circuits. Thus, if $\omega_1 = 1/\sqrt{L_1 C_1}$, $\omega_2 = 1/\sqrt{L_2 C_2}$ are the series-resonant angular frequencies and ω is the impressed angular frequency,

$$X_1 = \omega L_1 \left(1 - \frac{1}{\omega^2 L_1 C_1} \right) = \omega L_1 \left(1 - \frac{\omega_1^2}{\omega^2} \right) \quad (8.4)$$

$$X_2 = \omega L_2 \left(1 - \frac{1}{\omega^2 L_2 C_2} \right) = \omega L_2 \left(1 - \frac{\omega_2^2}{\omega^2} \right) \quad (8.5)$$

Instead of using the values of ω_1 and ω_2 to express the state of tuning of the two circuits, the series-resonant wavelengths λ_1 , λ_2 of the circuits can be used. By (4.2),

$$\frac{\omega_1}{\omega} = \frac{1}{\lambda_1/\lambda} \quad (8.6)$$

$$\frac{\omega_2}{\omega} = \frac{1}{\lambda_2/\lambda} \quad (8.7)$$

where λ is the wavelength in free space corresponding to ω .

Combining (8.4) and (8.6),

$$\frac{X_1}{\omega L_1} = 1 - \frac{\omega_1^2}{\omega^2} = 1 - \frac{1}{(\lambda_1/\lambda)^2} = \beta_1 \quad (8.8)$$

and, correspondingly,

$$\frac{X_2}{\omega L_2} = 1 - \frac{\omega_2^2}{\omega^2} = 1 - \frac{1}{(\lambda_2/\lambda)^2} = \beta_2 \quad (8.9)$$

The symbols β_1 and β_2 are introduced as abbreviations to denote the state of tuning of the primary and secondary circuits. They are

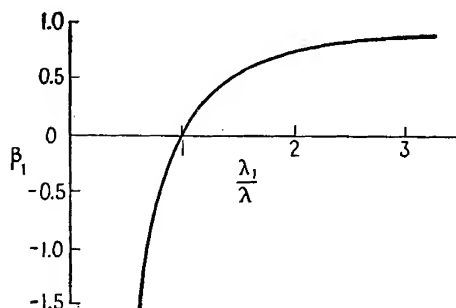


FIG. 8.1.—Tuning parameter β as a function of the ratio of series-resonant wavelength to wavelength of applied frequency. (Same for β_2 and λ_2/λ .)

dimensionless factors, and when they are multiplied by the reactance of the coil the product is the net reactance of the circuit containing this coil. The relation between β_1 and λ_1/λ is given in Fig. 8.1. The curve for β_1 applies also to β_2 .

Equation (8.1) now can be transformed into a simpler and better form for study by dividing numerator and denominator by $\omega^2 L_1 L_2$, giving

$$|I_2| = \frac{k|E_1|/\omega \sqrt{L_1 L_2}}{\sqrt{\left(\frac{1}{Q_1 Q_2} - \beta_1 \beta_2 + k^2\right)^2 + \left(\frac{\beta_2}{Q_1} + \frac{\beta_1}{Q_2}\right)^2}} \quad (8.10)$$

In this expression $k = |M|/\sqrt{L_1 L_2}$, $Q_1 = \omega L_1/R_1$, and $Q_2 = \omega L_2/R_2$.

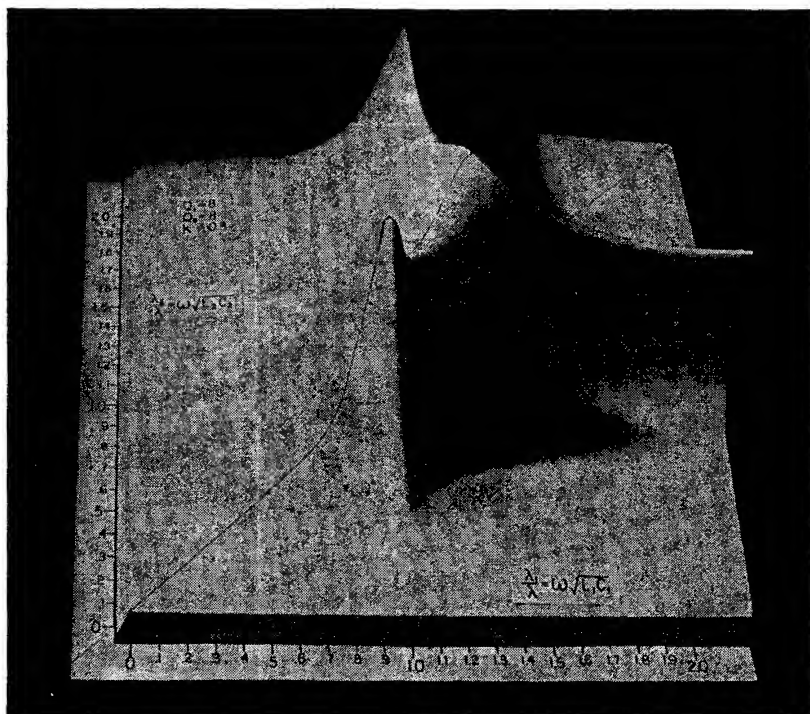


FIG. 9.1.—Space model of secondary current of two magnetically coupled circuits; $Q_1 = 8$, $Q_2 = 8$, $k = 0.3 > k_{\text{critical}}$.

9. Models for Secondary Current.—The manner in which $|I_2|$ changes as β_1 and β_2 are varied can be pictured by plotting $|I_2|$ vertically over a point on the horizontal plane determined by the coordinates β_1 and β_2 or λ_1/λ and λ_2/λ , thus forming a surface. For certain values of λ_1/λ and λ_2/λ , $|I_2|$ will have a maximum value denoted by $|I_2|_{\text{max}}$. Typical surfaces for $|I_2|/|I_2|_{\text{max}}$ are shown in Figs. 9.1 to 9.4. These figures differ as k , Q_1 , Q_2 are given various values.

Contour lines for various levels expressed as fractions of the maximum height are shown in Figs. 9.5 to 9.8, corresponding to Figs. 9.1 to 9.4.

The surfaces shown in the photographs and displayed by the contour charts are powerful aids in the understanding of the action of coupled circuits. They furnish a clearer picture of the depend-

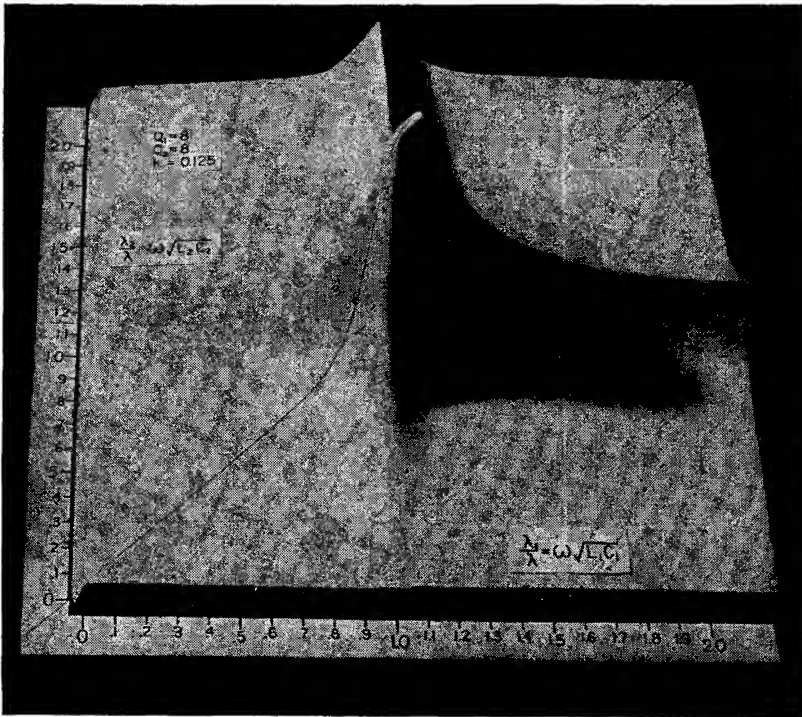


FIG. 9.2.—Space model of secondary current of two magnetically coupled circuits; $Q_1 = 8$, $Q_2 = 8$, $k = 0.125 = k_{\text{critical}}$.

ence of the secondary current upon the various factors than do the equations alone.

A general description of characteristics of the surfaces will be given first. In later sections these characteristics will be derived from the equations.

Consider first Fig. 9.5. For this case the coupling is somewhat large ($|M|$ is greater than in some of the other cases). The values of Q_1 and Q_2 are equal. This condition makes (8.10) symmetrical with respect to β_1 and β_2 , and the model is likewise sym-

metrical about a plane that makes a 45-degree angle with the reference axes. Also, the contour curves of Fig. 9.5 are symmetrical about a line making a 45-degree angle with the reference axis. There are two peaks, or maxima, of equal height, both lying on the 45-degree line. Between the two peaks the space model shows a saddle-shaped region.

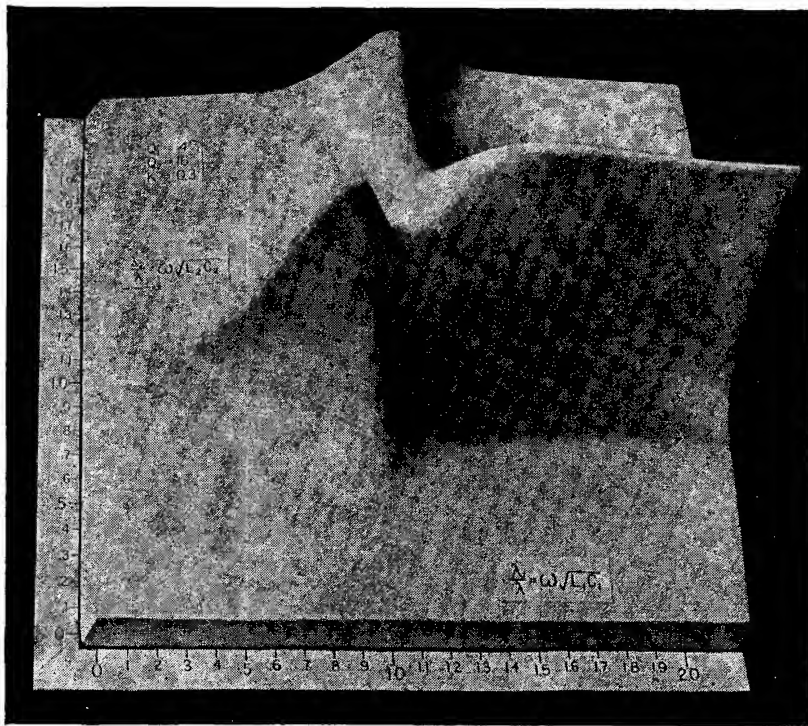


FIG. 9.3.—Space model of secondary current of two magnetically coupled circuits; $Q_1 = 4$, $Q_2 = 16$, $k = 0.3 > k_{\text{critical}}$.

Figure 9.6 is for the same conditions as Fig. 9.5 (equal Q 's) except that the coupling has been reduced to a value known as "critical" coupling (to be defined in Sec. 12). With this coupling there is only *one* peak, occurring where β_1 and β_2 are zero, or when both circuits are tuned to the frequency applied to the system.

For the surfaces of Figs. 9.7 and 9.8, the Q of the primary is different from that of the secondary, causing the asymmetry shown by the model. The coupling k is again large enough so that there are two peaks. The peaks are again of equal height, but they

are not located on the 45-degree line. As the coupling is reduced, the peaks approach each other, those of Fig. 9.8 being closer together than those of Fig. 9.7. If the coupling were reduced to the critical value, the two peaks would merge. For still weaker coupling, there would be a single peak of reduced height, located at the point corresponding to β_1 and β_2 equal to zero (no net reactance in either circuit). For still smaller values of k the single maximum

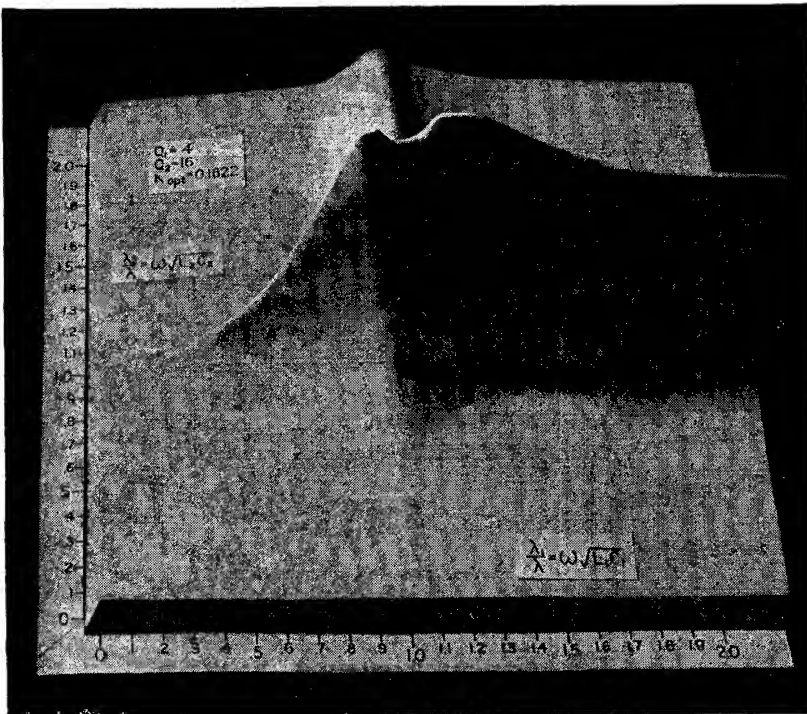


FIG. 9.4.—Space model of secondary current of two magnetically coupled circuits; $Q_1 = 4$, $Q_2 = 16$, $k = 0.1822 = k_{\text{optimum}}$.

of $|I_2|/|I_2|_{\text{max}}$ would decrease in height, remaining at the point for which β_1 and β_2 are zero.

It should be remarked here that the models in Figs. 9.1 to 9.4 are constructed for low values of Q in order to show clearly the separation of the peaks and to yield contour lines which are not too crowded. For higher values of Q the models and contour curves have the same general characteristics except that the peaks are sharper and confined to a smaller region on the base plane.

10. Conditions for Partial Resonance.—The conditions for and the value of the maximum secondary current at the peaks are of

considerable practical interest. However, it is necessary first to find what may be termed the *partial-resonance condition*. To do this the value of β_1 is held constant and the value of β_2 is found that gives a maximum value of $|I_2|$. Examination of (8.10) shows that, if the expression under the radical in the denominator is

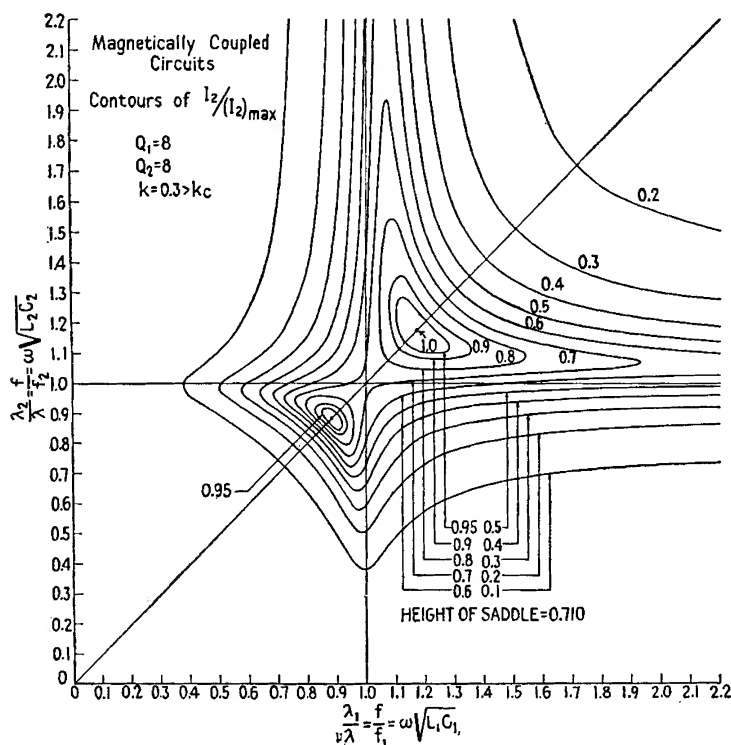


FIG. 9.5.—Contour curves of model of Fig. 9.1; $Q_1 = Q_2$, $k > k_{\text{critical}}$.

a minimum, the current $|I_2|$ is a maximum. To obtain this minimum value, differentiate with respect to β_2 the quantity under the radical (β_1 const.), and equate to zero, giving

$$2 \left(\frac{1}{Q_1 Q_2} - \beta_1 \beta_2 + k^2 \right) (-\beta_1) + 2 \left(\frac{\beta_2}{Q_1} + \frac{\beta_1}{Q_2} \right) \frac{1}{Q_1} = 0$$

whence

$$\beta_2 = \frac{k^2 \beta_1}{\frac{1}{Q_1^2} + \beta_1^2} \left] \begin{array}{l} \text{for partial resonance} \\ \text{primary fixed} \end{array} \right. \quad (10.1)$$

This equation states the value of β_2 for maximum $|I_2|$ for any chosen value of β_1 . This maximum is denoted by $\max |I_2|_{\beta_1}$ and is not necessarily equal to the greatest possible value of $|I_2|$ that may be obtained with the particular coupling employed. The greatest possible value of $|I_2|$ with the particular coupling employed

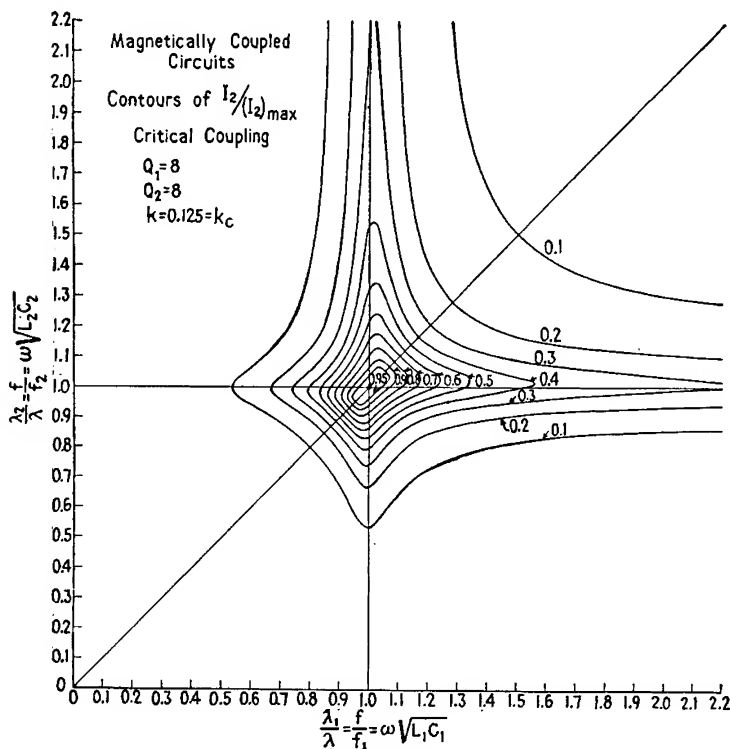


FIG. 9.6.—Contour curves of model of Fig. 9.2; $Q_1 = Q_2$, $k = k_{\text{critical}}$.

is denoted by $|I_2|_{\max}$ and is secured only when β_1 and β_2 have certain particular values, given in Sec. 11.

Equation (10.1) could have been obtained by inspection of the expression for I_2 as given by (6.9) if written in the equivalent form

$$I_2 = \frac{j\omega M E_1}{Z_{11} \left[R_2 + \frac{M^2 \omega^2 R_1}{R_1^2 + X_1^2} + j \left(X_2 - \frac{\omega^2 M^2 X_1}{R_1^2 + X_1^2} \right) \right]} \quad (10.2)$$

Since X_2 occurs only in the parentheses of the denominator, $|I_2|$ would be a maximum for any value of X_1 if

$$X_2 = \frac{\omega^2 M^2 X_1}{R_1^2 + X_1^2} \quad (10.3)$$

Equation (10.1) is the dimensionless form of this expression obtained by dividing (10.3) by ωL_2 . The locus of β_2 vs. β_1 , given by (10.1), is the locus of $\max |I_2|_{\beta_1}$ and is called the *partial-resonance primary-*

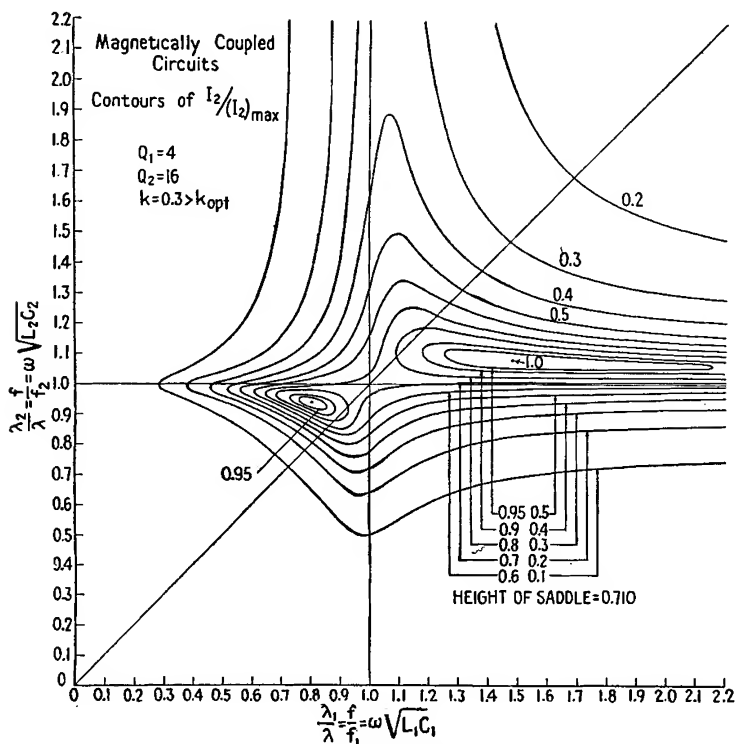


FIG. 9.7.—Contour curves of model of Fig. 9.3; $Q_1 < Q_2$, $k > k_{\text{optimum}}$.

fixed locus. This locus is plotted in Fig. 10.1 for the conditions of Fig. 9.7.

In similar manner the value of β_1 may be found to give a maximum $|I_2|$ for any given value of β_2 , when the tuning of the secondary circuit is first set and the primary circuit is then tuned for a maximum $|I_2|$, denoted by $\max |I_2|_{\beta_2}$. By symmetry, the condition is

$$\beta_1 = \frac{k^2 \beta_2}{\frac{1}{Q_2^2} + \beta_2^2} \quad \left. \vphantom{\beta_1} \right\} \begin{array}{l} \text{for partial resonance} \\ \text{secondary fixed} \end{array} \quad (10.4)$$

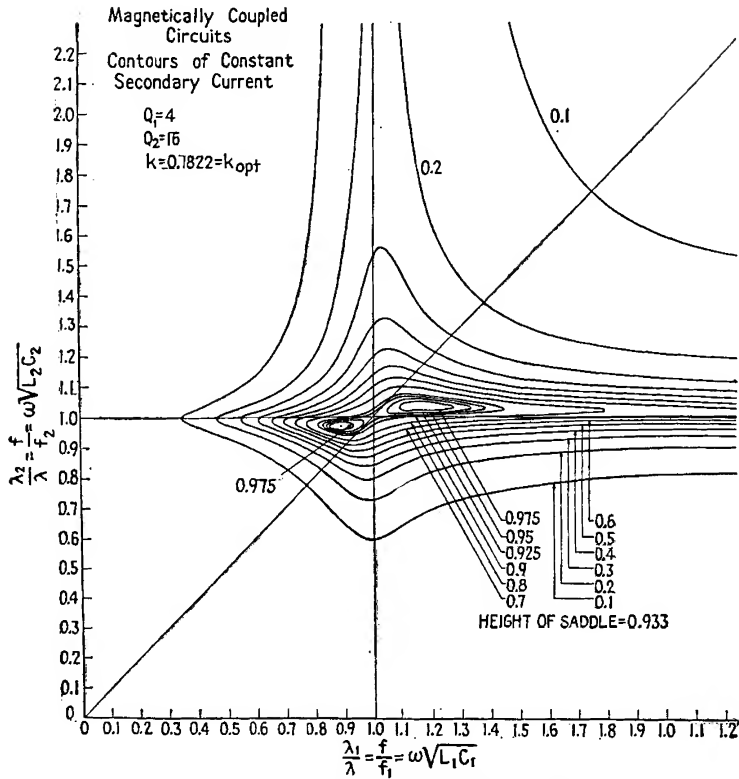


FIG. 9.8.—Contour curves of model of Fig. 9.4; $Q_1 < Q_2$, $k = k_{optimum}$.

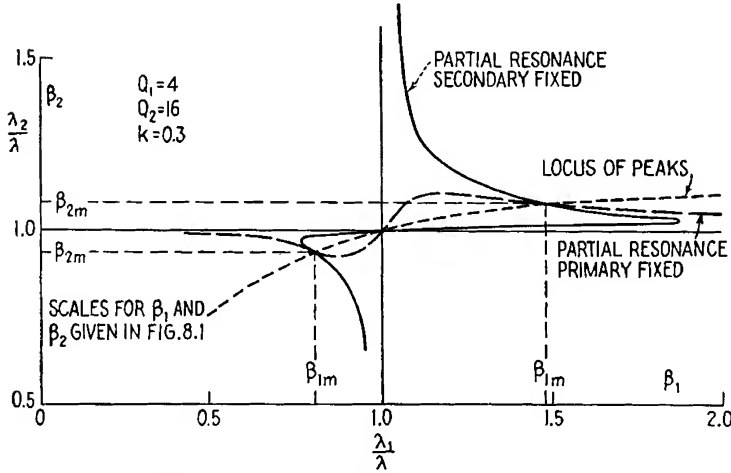


FIG. 10.1.—Partial-resonance conditions in magnetically coupled circuits.

11. Conditions for $|I_2|_{\max}$.—Equation (10.1) gives the value of β_2 for a max $|I_2|_{\beta_1}$ for any given value of β_1 . Surely, therefore, one value of β_1 , denoted by β_{1m} , and β_{2m} , the corresponding value of β_2 , must give the position of one of the main peaks of I_2 denoted by $|I_2|_{\max}$. The other main peak must lie above a point of the locus (10.1), as does the first peak.

Similarly, the two main peaks must also lie above the locus (10.4), the line *partial resonance secondary fixed*, of Fig. 10.1. Hence the main peaks, being on both loci, must be located at two of the intersection points of the two partial-resonance loci. The third intersection point, if there be three, cannot represent a maximum but does locate the lowest point in the “saddle-shaped” region between the main peaks.

The values of β_{1m} and β_{2m} for $|I_2|_{\max}$ can now be found by solving (10.1) and (10.4) simultaneously, which is the mathematical procedure for finding the points of intersection of two curves. Dividing (10.1) by (10.4),

$$\frac{\beta_{2m}^2}{\beta_{1m}^2} = \frac{\frac{1}{Q_2^2} + \beta_{2m}^2}{\frac{1}{Q_1^2} + \beta_{1m}^2} \quad (11.1)$$

Equation (11.1) can be simplified to

$$\frac{\beta_{2m}}{\beta_{1m}} = \frac{Q_1}{Q_2} \quad (11.2)$$

which describes the locus of the peaks. Since k does not appear in (11.2), this equation gives the relation between β_{1m} and β_{2m} for all values of k . If (11.2) is plotted as the dashed line in Fig. 10.1, it gives the path over which the peaks move as k is varied. If Q_1 is the same as Q_2 , this path is coincident with the 45-degree line.

Substituting in (10.4) the value of β_{2m} from (11.2) gives the values of β_{1m} ,

$$\beta_{1m} = \pm \sqrt{\frac{Q_2}{Q_1}} \sqrt{k^2 - \frac{1}{Q_1 Q_2}} \quad (11.3)$$

Substituting (11.3) or the value of β_{1m} from (11.2) in (10.1),

$$\beta_{2m} = \pm \sqrt{\frac{Q_1}{Q_2}} \sqrt{k^2 - \frac{1}{Q_1 Q_2}} \quad (11.4)$$

Equations (11.3) and (11.4) give the coordinates of $|I_2|_{\max}$, as in Fig. 10.1.

12. Critical Coupling.—As the coefficient of coupling k is reduced, the two peaks approach each other, moving along the dotted line in Fig. 10.1. The two peaks become one when the values of β_{1m} and β_{2m} become zero. The value of k that reduces (11.3) and (11.4) to zero is known as the *critical coupling* k_c , and

$$k_c = \frac{1}{\sqrt{Q_1 Q_2}} \quad (12.1)$$

If k is less than k_c , the values of β_{1m} and β_{2m} from (11.3) and (11.4) become imaginary and have no physical significance. For all values of k below critical coupling, there is only one peak, which is at $\beta_{1m} = 0$ and $\beta_{2m} = 0$.

Critical coupling may be defined also as follows: *Critical coupling* is the *greatest* coupling for which there is but *one* condition for a maximum of secondary current. *Critical coupling* is the coupling for which the secondary current has its maximum possible value for the condition that both circuits are tuned to the impressed frequency.

13. Value of $|I_2|_{\max}$.—If k is equal to or greater than critical coupling, the value of $|I_2|_{\max}$ for each peak can be found by substituting β_{1m} and β_{2m} from (11.3) and (11.4) into (8.10), whence

$$\begin{aligned} |I_2|_{\max} &= \left[\frac{|E_1| \sqrt{Q_1 Q_2}}{2\omega \sqrt{L_1 L_2}} \right]_{k \geq k_c} \\ &= \left[\frac{|E_1|}{2 \sqrt{R_1 R_2}} \right]_{k \geq k_c} \end{aligned} \quad (13.1)$$

Equation (13.1) shows that the height of both peaks is the same and remains constant as long as there are two peaks. That is, the maximum possible secondary current is independent of coupling k when k is greater than k_c . Also, (13.1) gives the greatest value that $|I_2|_{\max}$ can have and is the value corresponding to the peaks of Figs. 9.1 to 9.4 and indicated by the contour lines of Figs. 9.5 to 9.8.

To find the value of $|I_2|_{\max}$ for k less than k_c , β_1 and β_2 in (8.10) are given the value zero, whence

$$|I_2|_{\max} = \left[\frac{k|E_1|/\omega \sqrt{L_1 L_2}}{\frac{1}{Q_1 Q_2} + k^2} \right]_{k \leq k_c} \quad (13.2)$$

When $k \geq k_c$, $|I_2|_{\max}$ equals $E_1/(2 \sqrt{R_1 R_2})$. When $k < k_c$, $|I_2|_{\max}$ is less than $E_1/(2 \sqrt{R_1 R_2})$. A graph of (13.1) and (13.2) against k/k_c is shown in Fig. 13.1.

For $k > k_c$, the ratio of $|I_2|$ to $|I_2|_{\max}$ can be found by dividing (8.10) by (13.1). The ratio is

$$\frac{|I_2|}{|I_2|_{\max}} \Big|_{k \geq k_c} = \frac{2k \sqrt{Q_1 Q_2}}{\sqrt{[1 + Q_1 Q_2 (k^2 - \beta_1 \beta_2)]^2 + [\beta_1 Q_1 + \beta_2 Q_2]^2}} \quad (13.3)$$

This ratio is plotted in Figs. 9.1 to 9.8.

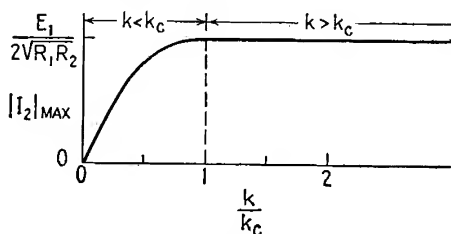


FIG. 13.1.—Variation of $|I_2|_{\max}$ with coefficient of coupling.

14. Recapitulation.—The purpose of this section is to recapitulate the important properties of magnetically coupled circuits given by the mathematical analysis of the preceding sections.

For each magnetically coupled system of two circuits there is a value k_c of the coefficient of coupling, known as critical coupling.

$$k_c = \frac{1}{\sqrt{Q_1 Q_2}} \quad (14.1)$$

For any coupling greater than critical coupling, the surface for $|I_2|$ has two maxima, or peaks, both of the same magnitude or height, given by

$$\begin{aligned} |I_2|_{\max} &= \frac{|E_1| \sqrt{Q_1 Q_2}}{2\omega \sqrt{L_1 L_2}} \Big|_{k \geq k_c} \\ &= \frac{|E_1|}{2 \sqrt{R_1 R_2}} \Big|_{k \geq k_c} \end{aligned} \quad (14.2)$$

If $Q_1 = Q_2$, these two maxima lie on the 45-degree line drawn on the plane having $\lambda_2/\lambda = f/f_2$ as ordinates and $\lambda_1/\lambda = f/f_1$ as abscissas, Fig. 9.5.

If Q_1 is not equal to Q_2 , these two maxima lie off the 45-degree line. The amount of detuning to reach a maximum is less for the circuit having the higher Q , as would be expected. The more common condition met in practice is that for which $Q_1 < Q_2$, as in Fig. 9.7.

As the coupling is decreased, the two peaks approach each other, moving toward the point $f/f_1 = 1$ and $f/f_2 = 1$, denoting

the condition that each circuit is tuned to the impressed signal. At critical coupling the two peaks merge into one, the maximum point lying at the point (1,1), as in Fig. 9.6 for $Q_1 = Q_2$.

For k less than k_c there is only one maximum, or peak, which remains at the point $f/f_1 = 1$ and $f/f_2 = 1$ but decreases in height as k decreases. Its value is given by

$$|I_2|_{\max} = \frac{k|E_1|/\omega \sqrt{L_1 L_2}}{\frac{1}{Q_1 Q_2} + k^2} \quad \left]_{k \leq k_c} \quad (14.3)$$

At critical coupling the ratio of the maximum voltage $|E_{c2}|_{\max}$ across the secondary capacitor to the series voltage $|E_1|$ in the primary circuit is important in intermediate-frequency amplifiers. Its value is

$$\left. \frac{|E_{c2}|_{\max}}{|E_1|} \right]_{k=k_c} = \frac{1}{2} \sqrt{\frac{L_2}{L_1}} \sqrt{Q_1 Q_2} \quad (14.4)$$

15. Variation of $|I_2|$ with ω .—A full knowledge of the manner in which $|I_2|$ varies with the frequency of the impressed signal

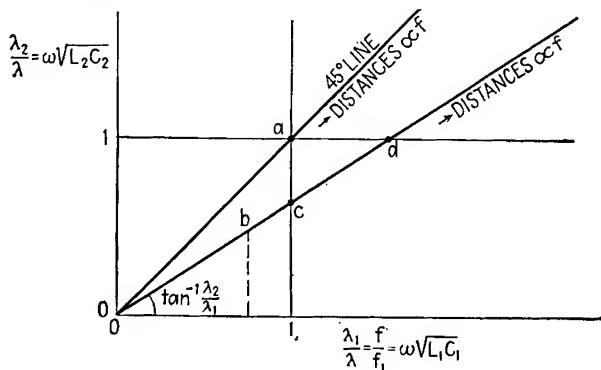


FIG. 15.1.—Indicating sections through coupled-circuit models and contour plots.

is of the greatest practical importance. A graph of $|I_2|$ vs. ω or f gives the equivalent resonance curve of the system and enables the band width of the system to be determined. One of the reasons for using coupled circuits in an intermediate-frequency amplifier is to obtain a more uniform amplification over a greater band width than can be obtained by an equivalent single circuit.

The models in Figs. 9.1 to 9.4 and the corresponding contour curves of Figs. 9.5 to 9.8 are of considerable value in visualizing the shape of the resonance curve with changing f . Suppose that

the circuits are tuned to the same frequency and that the impressed voltage E_1 has a frequency corresponding to the series-resonant frequency of the two circuits. The state of tuning would be represented by $\beta_1 = 0$ and $\beta_2 = 0$, determining a point on the base plane of the models, point *a*, Fig. 15.1. The coordinate axes of the models are marked off linearly in terms of $\lambda_1/\lambda = f/f_1 = \omega \sqrt{L_1 C_1}$ and $\lambda_2/\lambda = f/f_2 = \omega \sqrt{L_2 C_2}$. If now λ or f is varied, the two

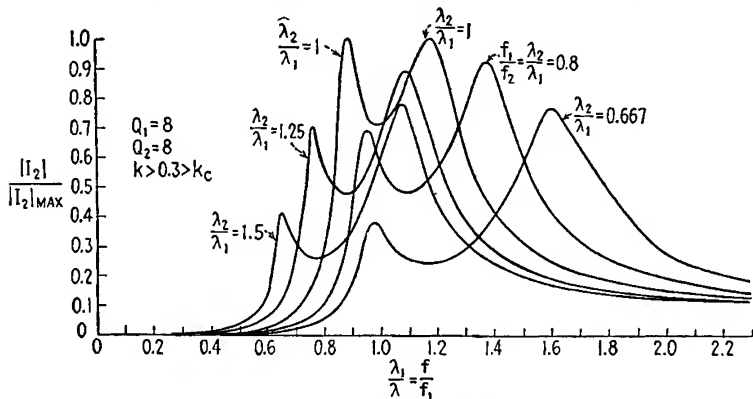


FIG. 15.2.—Sections of Fig. 9.5.

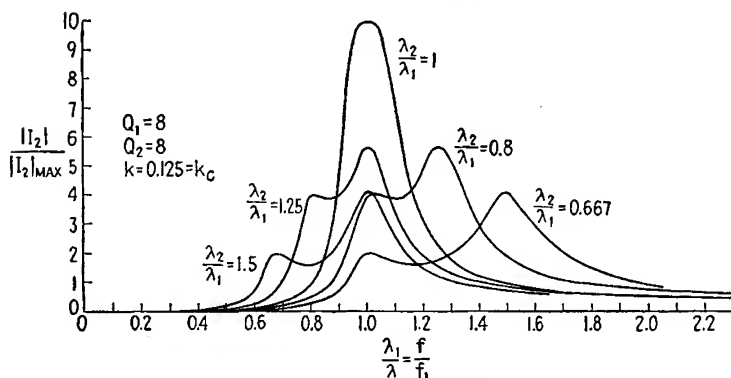


FIG. 15.3.—Sections of Fig. 9.6.

coordinates of the point vary proportionately and the point representing the state of tuning moves along the 45-degree line in Fig. 15.1. Distances along the 45-degree line, measured from the origin, are directly proportional to the frequency f of the impressed signal.

A cross section along the 45-degree line of any one of the models gives the variation of $|I_2|$ as f is varied, when both circuits are tuned to the same frequency.

If the circuits are not tuned to the same wavelength or frequency, a cross section of the models along a straight line passing through the origin and making an angle whose tangent is λ_2/λ_1 with the horizontal gives the variation of $|I_2|$ as f is varied. Such a section is indicated by the line $Obcd$ in Fig. 15.1. The frequency f for any point b , for example, is determined best by projecting b on the horizontal axis to find the value of f/f_1 . At point c the impressed frequency is that which would resonate the primary circuit alone; at point d the impressed frequency is that which would resonate the secondary circuit alone but would be about 50 per cent greater than the resonance frequency f_1 for the primary circuit. For

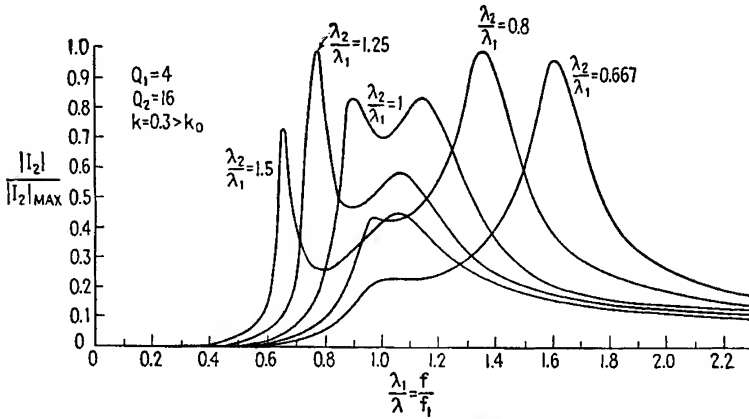


FIG. 15.4.—Sections of Fig. 9.7.

example, sections of the model of Fig. 9.1 along five radial lines for which λ_2/λ_1 or f_1/f_2 has the values 0.667, 0.8, 1, 1.25, 1.5 are shown in Fig. 15.2. Similar sections of the model of Fig. 9.2 for critical coupling are shown in Fig. 15.3. Sections for Fig. 9.7 where k is greater than k_0 and Q_2 is greater than Q_1 are shown in Fig. 15.4.

16. Resonance Curve when $f_1 = f_2$.—The proper tuning condition for an intermediate-frequency coupling system is that $f_2 = f_1$; hence the resonance curve is that given by a section of the models along the 45-degree line. To find this resonance curve put $\beta_2 = \beta_1$ in (13.3), repeated here in (16.1).

$$\left[\frac{|I_2|}{|I_2|_{\max}} \right]_{k \geq k_c} = \frac{2k \sqrt{Q_1 Q_2}}{\sqrt{[1 + Q_1 Q_2 (k^2 - \beta_1^2)]^2 + \beta_1^2 (Q_1 + Q_2)^2}} \quad (16.1)$$

Remember that $|I_2|_{\max}$ is the maximum value of the secondary

current attainable for a given coupling and may not be located on the 45-degree section.

Equation (16.1), plotted for $Q_1 = Q_2$ and for a coupling greater than critical coupling, is shown by the curve marked $\lambda_2/\lambda_1 = 1$ in Fig. 15.2. The curve for $Q_1 < Q_2$ is shown by the curve marked $\lambda_2/\lambda_1 = 1$ in Fig. 15.4. In both cases there are two maxima of equal height, with a minimum between at point $f/f_1 = 1$.

The positions of these maxima and minima on the f/f_1 scale can be found by differentiating (16.1) with respect to f/f_1 and equating the result to zero. Since Q_1 and Q_2 as well as β_1 are dependent upon f , this differentiation would be somewhat involved. But it should be noted that Q_1 and Q_2 vary much more slowly than β_1 when ω or f is varied, because β_1 is a difference of two terms that have about the same value. Furthermore, because of skin effect, the resistance of a circuit increases with frequency so that Q is generally more nearly constant than R . Neglecting the variation of Q with ω , the differentiation can be carried out with respect to β_1 ; and since β_1 occurs only in the denominator, the result can be obtained by differentiating the expression under the radical and equating the result to zero. The result is

$$\beta_1 = 0 \quad (16.2)$$

$$\beta_1 = \pm \sqrt{k^2 - \frac{1}{2} \left(\frac{1}{Q_1^2} + \frac{1}{Q_2^2} \right)} \quad (16.3)$$

The position of the minimum between the two maxima is given by (16.2). The two maxima are located at the two numerically equal values of β_1 given by (16.3). As k is decreased, the values of β_1 given by (16.3) decrease and the separation of the maxima decreases. When β_1 given by (16.3) becomes zero, the two maxima merge into a single maximum at $\beta_1 = 0$ or $f/f_1 = 1$. Hence, the largest value of k that gives a single maximum of I_2 along the 45-degree line is the value that makes β_1 of (16.3) equal to zero. This coupling is called *optimum coupling*,¹ denoted by k_o .

$$k_o = \sqrt{\frac{1}{2} \left(\frac{1}{Q_1^2} + \frac{1}{Q_2^2} \right)} \quad (16.4)$$

Optimum coupling is always greater than critical coupling when Q_1 is not equal to Q_2 but is identical with critical coupling when $Q_1 = Q_2$.

¹ In British literature optimum coupling is sometimes called *transitional coupling*.

The model in Fig. 9.4 and its contours in Fig. 9.8 are for optimum coupling. There are two peaks, located off the 45-degree line, showing that the coupling is greater than critical coupling. Follow-

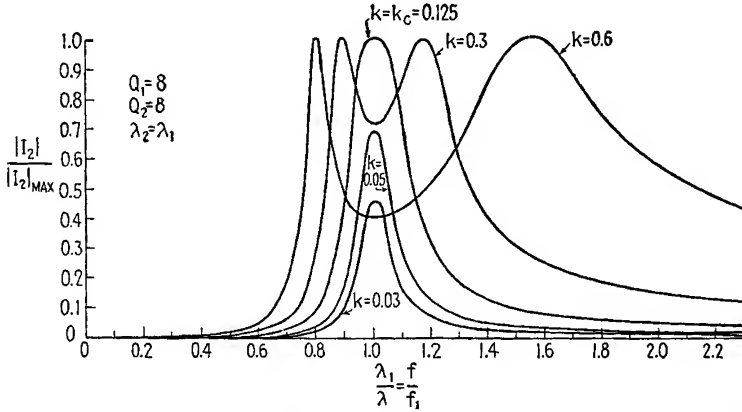


FIG. 16.1.—Variations in secondary current; circuits tuned to the same frequency and $Q_1 = Q_2$.

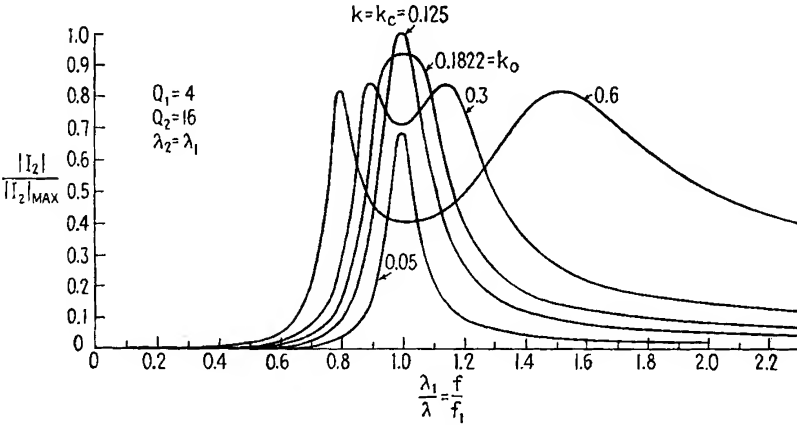


FIG. 16.2.—Variations in secondary current; circuits tuned to the same frequency and $Q_1 < Q_2$.

ing along the 45-degree line in Fig. 9.8 it is clear that there is only one maximum.

Resonance curves for various values of k when $\lambda_2 = \lambda_1$ and when $Q_1 = Q_2 = 8$ are shown in Fig. 16.1. The curve for $k = k_c = 0.125$ is a section of the model of Fig. 9.6. Note that, with equal Q 's, all maxima for $k \geq k_c$ have the same height, equal to $|E_1|/(2\sqrt{R_1 R_2})$.

Corresponding resonance curves for $\lambda_2 = \lambda_1$ but where $Q_1 < Q_2$ are shown in Fig. 16.2. The values of Q_1 and Q_2 are the same as those for Figs. 9.7 and 9.8. Two of the curves (*i.e.*, those for $k = 0.3$ and 0.1822) are sections of the models for these figures. The maximum of the resonance curve for k_o is less than the maximum for $k = k_o$.

17. Value of $|I_2|/|I_2|_{\max}$ for Optimum Coupling.—The value of $|I_2|/|I_2|_{\max}$ at the resonant frequency $f = f_1 = f_2$ for optimum coupling can be found by substituting in (16.1) the values $\beta_1 = 0$ and k_o from (16.4). The result is

$$\left. \frac{|I_2|}{|I_2|_{\max}} \right|_{\substack{k=k_o \\ f=f_1=f_2}} = \frac{2k_o \sqrt{Q_1 Q_2}}{1 + k_o^2 Q_1 Q_2} = \frac{2\sqrt{2} \sqrt{Q_2/Q_1} \sqrt{1 + (Q_2/Q_1)^2}}{\left(1 + \frac{Q_2}{Q_1}\right)^2} \quad (17.1)$$

If $Q_2 = Q_1$, (17.1) reduces to unity.

18. Band Width for Optimum Coupling.—The condition of optimum coupling provides the best single-peaked band-pass

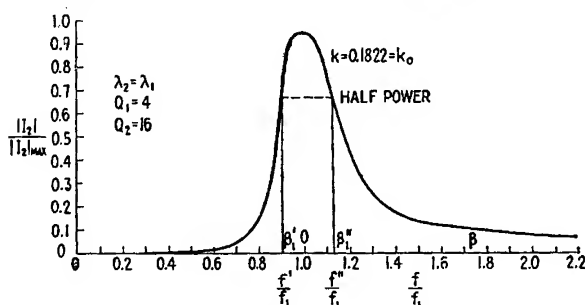


FIG. 18.1.—Variations in secondary current at optimum coupling.

resonance curve and is therefore the best condition for a selective intermediate-frequency amplifier that is required to pass a modulated signal. The band width of such a resonance curve is, therefore, of practical interest.

The fractional band width is defined as the frequency width of the curve at the half-power level divided by the frequency at the maximum of the curve. The ratio of the current at any point on the optimum-coupling resonance curve for $f_2 = f_1$ to the current at the peak of this resonance curve can be found by dividing (16.1) with $k = k_o$ by (17.1), giving

$$\left. \frac{|I_2|}{|I_2|_{f=f_1=f_2}} \right|_{f=f_2} = \frac{1 + k_o^2 Q_1 Q_2}{\sqrt{[1 + Q_1 Q_2 (k_o^2 - \beta_1^2)]^2 + \beta_1^2 (Q_1 + Q_2)^2}} \quad (18.1)$$

The next step is to find the two values of β_1 that make the left-hand side of (18.1) equal to $1/\sqrt{2}$. The result is

$$\beta_1 = \pm \frac{1}{\sqrt{2}} \left(\frac{1}{Q_1} + \frac{1}{Q_2} \right) \quad (18.2)$$

These two values of β_1 may be called β_1' and β_1'' as in Fig. 18.1, which is a graph of (18.1) for $Q_1 = 4$ and $Q_2 = 16$. Then expanding β_1' and β_1'' gives

$$\frac{f''}{f_1} = \frac{1}{\sqrt{1 - \beta_1''}} \quad \text{and} \quad \frac{f'}{f_1} = \frac{1}{\sqrt{1 - \beta_1'}} \quad (18.3)$$

Hence the band width is

$$\frac{f'' - f'}{f_1} = \frac{1}{\sqrt{1 - \beta_1''}} - \frac{1}{\sqrt{1 - \beta_1'}} \doteq \frac{1}{2}(\beta_1'' - \beta_1') \quad (18.4)$$

This gives as, an approximate value for band width,

$$\text{Band width for } k_o = \frac{f'' - f'}{f_1} \doteq \frac{1}{\sqrt{2}} \left(\frac{1}{Q_1} + \frac{1}{Q_2} \right) \quad (18.5)$$

The band width for critical coupling can be obtained in a similar manner and is given by the following approximate formula:

$$\text{Band width for } k_c = \frac{1}{\sqrt{2}} \sqrt{\sqrt{\left(\frac{1}{Q_1} - \frac{1}{Q_2} \right)^4 + \frac{16}{Q_1^2 Q_2^2}} - \left(\frac{1}{Q_1} - \frac{1}{Q_2} \right)^2} \quad (18.6)$$

19. Band-width Curves for k_c and k_o .—Curves of constant band width for various values of Q_1 and Q_2 are plotted in Fig. 19.1. The curves of constant band width for optimum coupling k_o are obtained from the exact form of (18.4) by plotting Q_1 against Q_2 for constant values of $(f'' - f')/f_1$. The curves of constant band width for critical coupling k_c are derived from (18.6). The 45-degree lines with positive slopes are lines over which Q_2/Q_1 have constant values as indicated on the lines.

The manner of using these curves is explained by the following examples: If $Q_2 = Q_1$, the data must be read off the 45-degree line marked $Q_2/Q_1 = 1$. For a band width of 0.01 with equal Q 's, Q_1 and Q_2 must be 141; optimum coupling is the same as critical coupling. Suppose now that Q_2 is $10Q_1$ and that Q_1 is 21.6, as

indicated by point *a* in Fig. 19.1. The band width with critical coupling would be 0.01 but with optimum coupling would be 0.036.

The spreading apart of the lines for optimum and critical coupling shows that, as the ratio of Q_2/Q_1 departs from unity, optimum coupling gives a greater and greater band width as compared with that given by critical coupling.

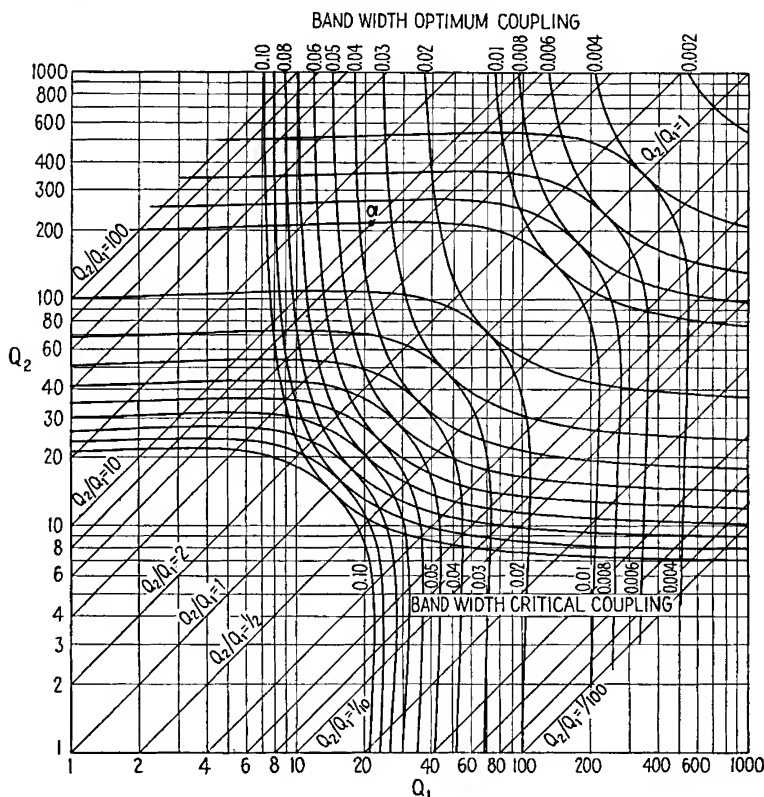


FIG. 19.1.—Fractional band width for critical and optimum coupling of two magnetically coupled circuits.

The curves also show that, if Q_2 is at least ten times Q_1 , the band width is practically independent of Q_2 .

20. Parallel-fed Coupled Systems.—In the preceding discussion the power has been fed to the coupled circuits from a series source denoted by E_1 in Figs. 6.1 and 6.2. Frequently, however, the a-c power is fed in parallel with the primary circuit, the primary circuit being in the plate circuit of a vacuum tube as in Fig. 20.1*a*. The subscripts 1 and 2 indicating primary and secondary circuits are

replaced by subscripts b and c in order to conform with the notation used in vacuum-tube circuits.

The vacuum tube operating as a linear Class A amplifier¹ may be replaced by a generator of voltage μE_g in series with the plate resistance r_p . Instead of treating this case as a new problem it is easier to transform the circuit into an equivalent series circuit and then to apply the coupled-circuit theory that has been developed

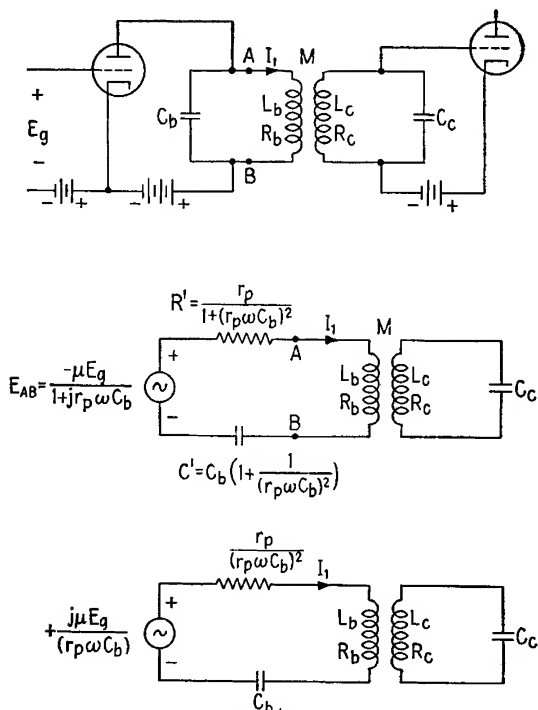


FIG. 20.1.—Parallel-fed coupled system and its equivalent circuit.

in the preceding sections. This transformation can be made by applying Thévenin's theorem to the circuit of Fig. 20.1a at points AB as explained in Chap. VI, Sec. 5. According to Thévenin's theorem the portion of the circuit to the left of points AB can be replaced by a series generator having a voltage E_{AB} equal to the open-circuit voltage between points A and B , in series with an impedance Z_{AB} equal to the impedance between points A and B with the voltage suppressed.

¹ See Chap. XIII.

The open-circuit voltage E_{AB} is given by (5.1), Chap. V. The impedance Z_{AB} is given by (5.2), Chap. V. Then the circuit of Fig. 20.1a can be transformed to the circuit in Fig. 20.1b. If, as is often the case, the dimensionless quantity $(r_p\omega C_b)^2$ is very large in comparison with unity, the equivalent series circuit assumes the form in Fig. 20.1c. In this case C_b is the same as for the parallel-fed circuit, but the added resistance due to r_p is $r_p/(r_p\omega C_b)^2$ hence is much smaller than r_p . The series voltage is also advanced by 90° and is equal in magnitude to $\mu E_g/r_p\omega C_b$.

The effective Q_1 of the primary circuit is

$$Q_1 \doteq \frac{\omega L_b}{R_b + \frac{r_p}{(r_p\omega C_b)^2}} \quad (20.1)$$

The added resistance $r_p/(r_p\omega C_b)^2$ generally makes Q_1 less than Q_2 .

21. Power Transfer in Coupled Circuits.—In amplifier and receiving circuits, voltage gain and band width are generally the

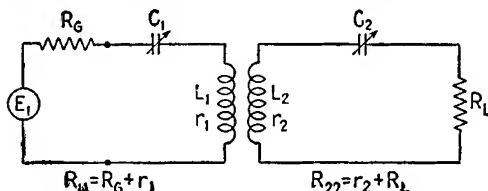


FIG. 21.1.—Magnetically coupled system.

most important quantities to be considered in the adjustments of the circuits. In dealing with systems such as transmitters the efficiency of power transfer from the generator to the load is an important consideration. Often, coupled circuits are used in such systems, and the conditions for the most efficient transfer of power through the circuits is worthy of some study.

Commonly in power systems a generator of some sort feeds power into one of two or more coupled oscillatory circuits, and the load is connected in one of the other oscillatory circuits. This arrangement is illustrated in Fig. 21.1 where r_2 is the resistance of the secondary circuit exclusive of the load resistance R_L and R_{22} is the total resistance of the secondary circuit. Similarly, r_1 is the circuit resistance of the primary circuit exclusive of the generator resistance R_g , and R_{11} is the sum of r_1 and R_g and is the total resistance of the primary circuit.

The object is to transfer power from E_1 to R_L , but the power in R_L is proportional to the power transferred to the secondary circuit.

The total efficiency η_t of transfer of power may be written

$$\eta_t = \eta_{cir}\eta_2 \quad (21.1)$$

where η_{cir} is the circuit efficiency and represents the efficiency of transfer of power from E_1 to the secondary circuit and η_2 represents the proportion of useful power (in R_L) to the total power in the secondary circuit and is equal to R_L/R_{22} . The present study is concerned only with η_{cir} .

The circuit efficiency η_{cir} is the power dissipated in R_{22} divided by the total power input, which is the power dissipated in both R_{11} and R_{22} , or

$$\eta_{cir} = \frac{I_2^2 R_{22}}{I_1^2 R_{11} + I_2^2 R_{22}} = \frac{1}{\frac{I_1^2 R_{11}}{I_2^2 R_{22}} + 1} \quad (21.2)$$

where I_1 and I_2 are magnitude values. But (6.2) gives, for magnitude values,

$$\frac{I_1^2}{I_2^2} = \frac{Z_{22}^2}{Z_{12}^2} = \frac{R_{22}^2 + X_{22}^2}{\omega^2 M^2} \quad (21.3)$$

Substituting (21.3) in (21.2),

$$\eta_{cir} = \frac{1}{\frac{R_{22}^2 + X_{22}^2}{\omega^2 M^2} \frac{R_{11}}{R_{22}} + 1} \quad (21.4)$$

Equation (21.4) shows that the efficiency η_{cir} is not dependent upon the value of X_{11} but does depend upon X_{22} and is a maximum when X_{22} is zero. This important result is restated as follows:

For maximum efficiency of power transfer in magnetically coupled circuits, the circuit to which power is transferred should be tuned to resonance.

In the more general case in Fig. 6.1, if R_{12} is negligible, maximum efficiency of power transfer results when $X_{22} = 0$. Setting $X_{22} = 0$ in (21.4) gives as maximum efficiency

$$\max \eta_{cir} = \frac{1}{\frac{R_{11}R_{22}}{\omega^2 M^2} + 1} \quad (21.5)$$

$$= \frac{k^2 Q_1 Q_2}{1 + k^2 Q_1 Q_2} \quad (21.6)$$

where Q_1 and Q_2 are now given by $Q_1 = \omega L_1/R_{11}$ and $Q_2 = \omega L_2/R_{22}$. Using (14.1), $Q_1 Q_2$ may be replaced by $1/k_c^2$, and

$$\max \eta_{cir} = \frac{(k/k_c)^2}{1 + (k/k_c)^2} \quad (21.7)$$

Equation (21.7) shows that when $k = k_c$ the efficiency is 50 per cent, and to obtain higher efficiencies k must be greater than k_c . A ratio k/k_c of 5 gives an efficiency of approximately 96 per cent. Hence k/k_c should be within the range from 3 to 6.

When the secondary is tuned to resonance, $X_{22} = 0$, and the impedance of the coupled system of Fig. 21.1 presented to E_1 is, from (7.2),

$$Z_{11}' \Big|_{X_{22}=0} = R_{11} + \frac{\omega^2 M^2}{R_{22}} + jX_{11} \quad (21.8)$$

$$= R_{11} \left[1 + \left(\frac{k}{k_c} \right)^2 \right] + jX_{11} \quad (21.9)$$

If X_{11} is zero, the coupled system presents to E_1 a load of unity power factor. The setting of the circuits is at $f/f_1 = 1$ and $f/f_2 = 1$ on the models in Figs. 9.1 and 9.3, or directly on the low point of the saddle. The adjustment to give $|I_2|_{\max}$ is not the condition of greatest efficiency except when the coupling is equal to or less than critical coupling.

If the primary circuit be considered as the impedance of the generator and the secondary circuit be the load circuit, the condition for maximum power transfer is that $\omega^2 M^2 / Z_{22}$ be the conjugate of Z_{11} , that is,

$$R_{11} = \frac{\omega^2 M^2 R_{22}}{R_{22}^2 + X_{22}^2} \quad (21.10)$$

$$X_{11} = \frac{\omega^2 M^2 X_{22}}{R_{22}^2 + X_{22}^2} \quad \text{or} \quad \beta_1 = \frac{\beta_2 k^2}{\frac{1}{Q_2^2} + \beta_2^2} \Bigg]_{\text{for max. power transfer}} \quad (21.11)$$

where $\beta_1 = X_{11}/\omega L_1$ and $\beta_2 = X_{22}/\omega L_2$. These are the conditions for $|I_2|_{\max}$ since (21.11) is the same as (10.4) and (21.10) divided by (21.11) gives (11.2). But substituting (21.10) in (21.4) gives an efficiency of 50 per cent. This shows that whenever the circuits are adjusted to give $|I_2|_{\max}$ for $k \geq k_c$ the circuit efficiency is 50 per cent.

The preceding demonstration has shown that, when the circuit is driven by a generator of constant internal impedance, the condition of maximum power transfer to the secondary circuit and the condition for maximum efficiency are not identical.

CHAPTER VIII

FILTERS

1. Definitions.—An *ideal low-pass filter* would transmit without loss alternating power of any frequency lower than a given limit f_c , called the *cutoff* frequency, and would block altogether the transmission of alternating power of frequency higher than the cutoff frequency. The frequency band from 0 to f_c is called the *transmission band*; the band from f_c to ∞ is called the *attenuation band*. A *high-pass filter* has 0 to f_c for the attenuation band and f_c to ∞ for the transmission band. A *single band-pass filter* has a finite frequency band f_1 to f_2 for the transmission band; 0 to f_1 and f_2 to ∞ are attenuation bands. A *single band-stop filter* has 0 to f_1



FIG. 1.1.—Block diagram of filtering circuit.

and f_2 to ∞ for transmission bands and f_1 to f_2 for the attenuation band.

A block diagram of a filtering circuit inserted between a generator and a load is shown in Fig. 1.1. If the filter contains resistance, it will dissipate some energy. A filter should preferably have no losses, *i.e.*, no dissipation. This cannot be attained in practice but can be approached closely. The development and practical application of the formulas used in filter design are greatly facilitated by considering the dissipationless case. Therefore, the theory of the dissipationless filter must be regarded as a description of conditions that are approached in the limit, as the dissipation of the filter is made less and less and finally becomes zero. In practice, the filter elements are designed with as high a Q (ratio of reactance to resistance) as is economically possible over the range of frequencies used.

2. Low-pass Filter.—A circuit containing series inductors and shunt capacitors will pass low frequencies and attenuate high frequencies. The simplest circuit to analyze is one having two

equal series coils without mutual inductance. In practice, the load is commonly a pure resistance R , constant over the range of frequencies used. To secure maximum power transfer, the generator should have an internal impedance R_g equal to R , Fig. 2.1. The filter is inserted between the generator and the load, Fig. 2.2. It is desirable that the filter terminated by the load present the same impedance R to the generator as the load itself. It will be shown in Sec. 4 that this is approximately true when

$$\sqrt{\frac{L}{C}} = R \quad (2.1)$$

The derivations that follow in this section are based upon this condition.

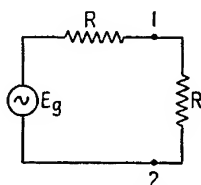


FIG. 2.1.—Generator and load matched for maximum power transfer.

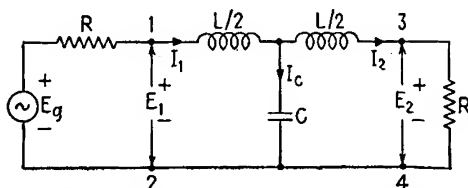


FIG. 2.2.—Filtering network introduced between generator and load of Fig. 2.1.

It is possible to obtain the voltage E_2 across the load, Fig. 2.2, from the Kirchhoff equations of the network. The derivations are rather long and, not being essential, are not given here. The magnitude of E_2 is found to be

$$|E_2| = \frac{|E_g|}{2} \frac{1}{\sqrt{1+x^6}} \quad (2.2)$$

and E_2 lags E_g by β radians, β being given by

$$\tan \beta = \frac{2x - x^3}{1 - 2x^2} \quad (2.3)$$

In these formulas the numerical factor x is given by

$$x = \pi \sqrt{LC} f = \frac{f}{f_c} \quad (2.4)$$

where f is the applied frequency and $f_c = 1/(\pi \sqrt{LC})$.

Figure 2.3 shows $|E_2|$ and β as functions of frequency. The ratio of $|E_2|$ to $|E_g|/2$ is practically unity for all frequencies below

$f_c/2$; for higher frequencies the ratio decreases gradually to zero. The phase lag β increases from zero to $3\pi/2$ (or 270°) and is $3\pi/4$ (or 135°) at f_c .

Without the filter, the power delivered to the load would be $P_0 = E_g^2/4R$, with the filter it is $P_f = I_2 E_2 = E_2^2/R$. The power ratio

$$\frac{P_0}{P_f} = \left(\frac{E_g}{2E_2} \right)^2 = 1 + x^6 \quad (2.5)$$

is a measure of the effectiveness of the filter. The number

$$D \text{ (db)} = 10 \log \left(\frac{P_0}{P_f} \right) \quad (2.6)$$

is called the *insertion loss* (in decibels) due to the filter. In this case, the loss is practically zero for all frequencies below $f_c/2$,

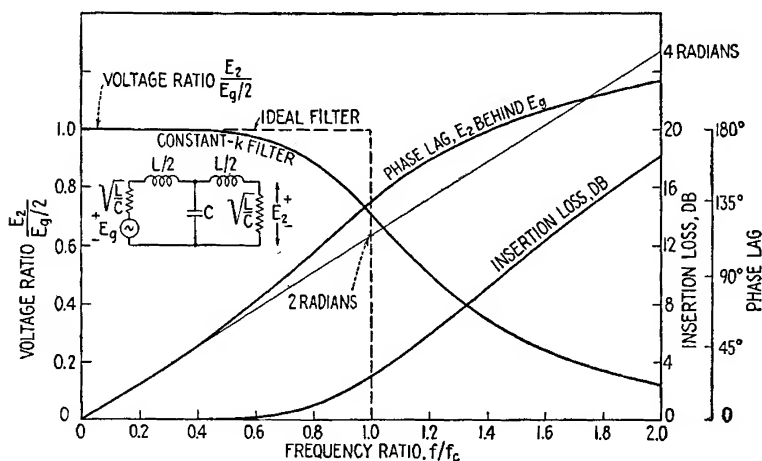


FIG. 2.3.—Frequency characteristics of constant- k low-pass network with fixed resistive load.

Fig. 2.3. It is 3 db (half power) at f_c and for higher frequencies increases indefinitely.

It is seen from Fig. 2.3 that the network just analyzed does not discriminate sharply between frequencies above and below a certain frequency f_c , as the *ideal* low-pass filter defined in Sec. 1 should do; rather, its response changes gradually from one end of the spectrum to the other. This network, however, may be considered as an *approximation* to an ideal filter whose cutoff would be f_c . The approximation is not very good but is as good as can be obtained with only two coils $L/2$ and one capacitor C .

3. Ladder-type Low-pass Filter.—A sharper frequency discrimination is obtained if two or more identical filter *sections* are connected in cascade between generator and load, Fig. 3.1. In such a circuit any two coils connected in series may be replaced by one coil only, Fig. 3.2. This form of network is called a *ladder-type filter* from the appearance of its circuit diagram. If n sections

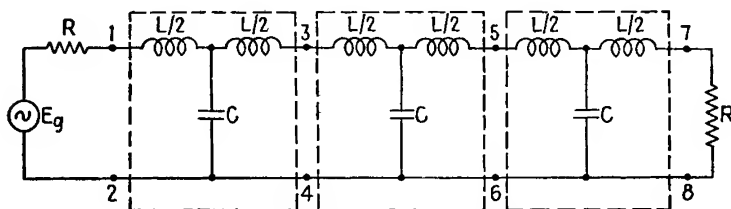


FIG. 3.1.—Three-section low-pass filter.

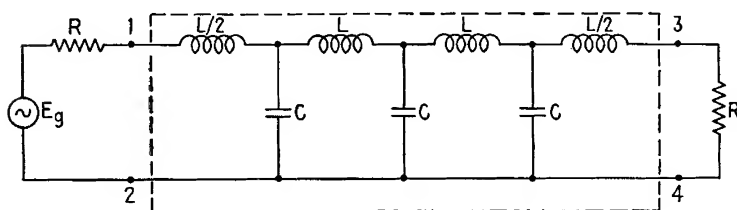


FIG. 3.2.—A ladder-type filter equivalent to the filter of Fig. 3.1.

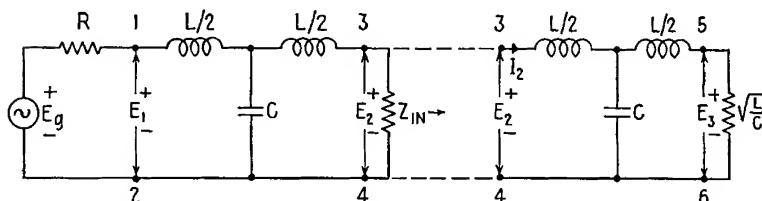


FIG. 3.3.—Calculation of the output voltage of a two-section filter.

are used, it consists of $n - 1$ identical coils L in series, plus a coil $L/2$ at each end, and n capacitors C in shunt.

Since a two- or three-section filter is more effective than a one-section filter, it is desirable to calculate the voltages and currents in such networks. Unfortunately the calculations become more and more complicated as the number of sections increases. The reason for this is already apparent in the case of $n = 2$. To calculate the input impedance at terminals 1-2, and then the input current, replace the second filter section, terminated by the constant resistance $R = \sqrt{L/C}$, by its input impedance Z_{in} , Fig. 3.3. The

first section is terminated by this impedance Z_{in} , which, however, is not a constant resistance but is a function of frequency given by

$$Z_{in} = R \frac{1 - 2j(x^3 - 2x^5)}{1 + 4x^4} \quad (3.1)$$

where $x = f/f_c$, and is plotted in Fig. 3.4. The input impedance of the two-section filter is a still more complicated function of frequency, and the complication increases with the number of sections.

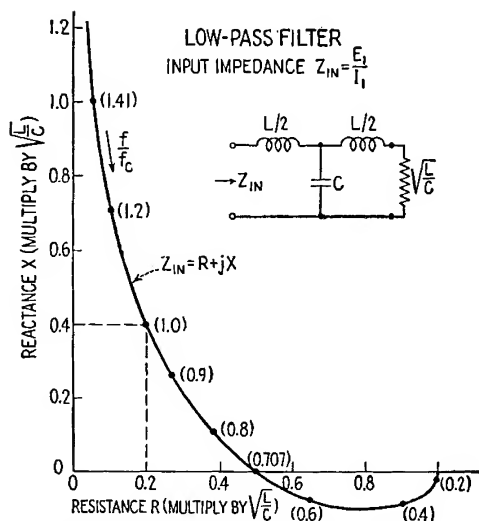


FIG. 3.4.—Input impedance of the filter network of Fig. 2.3, as function of the frequency ratio f/f_c .

Calculations, however, are greatly simplified if the termination, instead of being a constant resistance, is assumed to be a certain specified impedance. The general theory of filters deals with sections that are equivalent to a T with two unequal series arms. The present treatment will consider only symmetrical sections, for which the specified impedance is called the *characteristic impedance* of the symmetrical filter.

4. Characteristic Impedance of Low-pass Filter.—Let f be the applied frequency. By definition, the characteristic impedance Z_c at frequency f has such a value that, when the symmetrical one-section filter is terminated by Z_c , its input impedance measured at terminals 1-2 is equal to Z_c , Fig. 4.1.

The input impedance of the network, Fig. 4.1, is

$$Z_{in} = jX_{L/2} + \frac{jX_c(jX_{L/2} + Z_c)}{jX_c + jX_{L/2} + Z_c} \quad (4.1)$$

In terms of R and x , the reactance of each series coil is

$$X_{L/2} = j\omega \frac{L}{2} = j\sqrt{\frac{L}{C}} \frac{\sqrt{LC}}{2} \omega = jRx \quad (4.2)$$

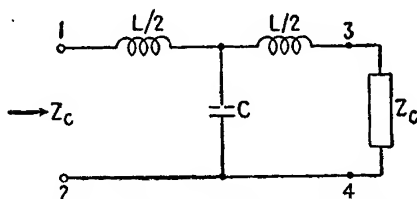


FIG. 4.1.—Characteristic impedance of low-pass filter.

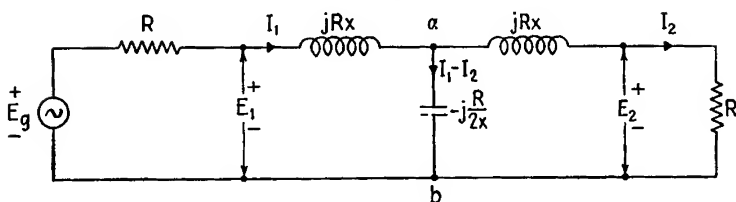


FIG. 4.2.—Network of Fig. 4.1, with values of the reactances inserted; $x = f/f_c$.

and the reactance of the shunt capacitor is

$$X_c = -j\frac{1}{\omega C} = -j\sqrt{\frac{L}{C}} \frac{2}{\sqrt{LC}} \frac{1}{2\omega} = -j\frac{R}{2x} \quad (4.3)$$

as indicated in Fig. 4.2. Replacing Z_{in} by Z_c in (4.1) and solving for Z_c , it is found that

$$Z_c = \sqrt{\frac{L}{C}} \sqrt{1 - \left(\frac{f}{f_c}\right)^2} = R \sqrt{1 - x^2} \quad (4.4)$$

Two cases must be distinguished:

CASE 1: $f < f_c$, or $x < 1$. For all frequencies below the cutoff, the characteristic impedance is a *pure resistance*, the value of which decreases from $R = \sqrt{L/C}$ for very low frequencies to zero for f_c , Fig. 4.3a.

CASE 2: $f > f_c$, or $x > 1$. In this case

$$Z_c = jR \sqrt{x^2 - 1} = jX \quad (4.5)$$

For all frequencies above the cutoff, the characteristic impedance is a *pure inductance*,¹ increasing from zero at f_c to $j\omega L/2$ at very high frequencies, Fig. 4.3b.

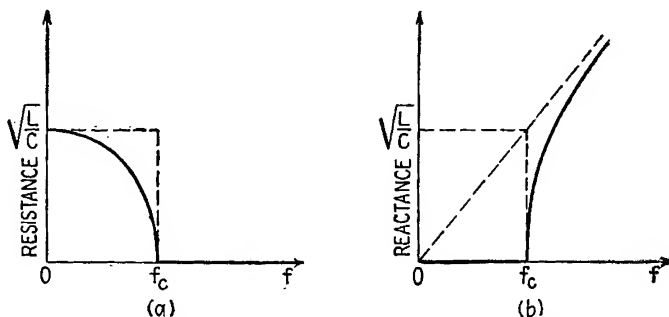


FIG. 4.3.—Variations with frequency of (a) the resistive part of Z_c and (b) the reactive part of Z_c .

The variations of Z_c plotted in the complex plane, Fig. 4.4, give two perpendicular straight lines. There is no actual combination of circuit elements that has exactly these variations in impedance. However, these variations are approximated by the input impedance of one filter section terminated in $R = \sqrt{L/C}$, as shown by Fig. 3.4 and by the dotted line of Fig. 4.4. This approximation becomes better and better the larger the number of filter sections connected in cascade. It is generally assumed, therefore, that filters are terminated by their characteristic impedance rather than by a constant resistive load, since this assumption simplifies calculations and yields results that are sufficiently accurate to be useful.

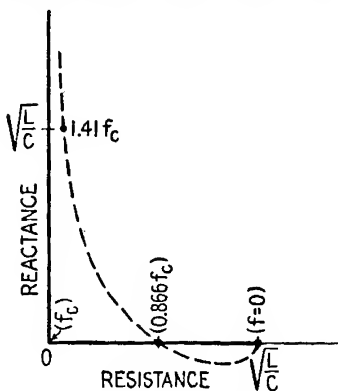


FIG. 4.4.—Variations with frequency of the characteristic impedance Z_c (heavy solid lines) of a one-section low-pass filter, and of the input impedance Z_{in} (dotted line) when the termination is $R = \sqrt{L/C}$ as in Fig. 3.4.

¹ In the general case of a dissipative low-pass filter, the Z_c from (4.4) could be in the attenuation band either $R + jX$ or $-R - jX$ (where R and X are positive quantities). The latter solution is usually disregarded since the practical application of filters is to dissipative loads. In the limit, as the filter becomes dissipationless, R decreases to zero, leaving $Z_c = jX$ as in (4.5).

5. Attenuation and Phase Lag of the Low-pass Filter.—When the filter is terminated in its characteristic impedance, Fig. 5.1, a simple relation holds between the input and output voltages and currents, *viz.*,

$$\frac{E_1}{I_1} = \frac{E_2}{I_2} = Z_c \quad (5.1)$$

since the input and output impedances are *both* equal to Z_c . From (5.1),

$$\frac{E_2}{E_1} = \frac{I_2}{I_1} \quad (5.2)$$

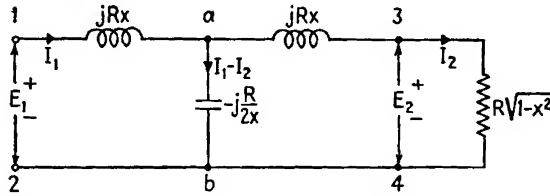


FIG. 5.1.—One-section low-pass filter terminated by its characteristic impedance.

This common ratio now will be calculated. The voltage across the capacitor, Fig. 5.1, is equal to the sum of the voltages across the right-hand coil and the load, or

$$-j \frac{R}{2x} (I_1 - I_2) = (jRx + R \sqrt{1 - x^2}) I_2 \quad (5.3)$$

whence

$$I_1 = [(1 - 2x^2) + 2jx \sqrt{1 - x^2}] I_2 \quad (5.4)$$

The quantity in brackets is in general a complex numeric, which has a magnitude and an angle. The magnitude, being a positive number, may be written as ϵ^α ; the angle may be called β . The physical meaning of the symbols α and β will appear shortly. The numeric then may be written as

$$\epsilon^\alpha (\cos \beta + j \sin \beta) = \epsilon^\alpha \epsilon^{j\beta} = \epsilon^{\alpha + j\beta} \quad (5.5)$$

so that

$$I_1 = \epsilon^{\alpha + j\beta} I_2 \quad (5.6)$$

It will be shown now that (5.6) corresponds to two completely different conditions, according to whether the applied frequency is below or above the cutoff.

CASE 1: $f < f_c$, or $x < 1$. In this case $(1 - 2x^2)$ and $2x \sqrt{1 - x^2}$ are two real quantities, the sum of whose squares is

unity. The magnitude of the complex quantity in (5.4) is then $1 = e^\alpha$; hence $\alpha = 0$, and, by comparing (5.4) and (5.5),

$$1 - 2x^2 = \cos \beta \quad 2x \sqrt{1 - x^2} = \sin \beta \quad (5.7)$$

The magnitudes $|I_1|$ and $|I_2|$ are equal, and I_2 lags I_1 by β radians. From (5.7) it is seen that β increases from 0 to π as x increases from 0 to 1, that is, as f increases from 0 to f_c . In this case, (5.6) may be written as

$$I_1 = e^{j\beta} I_2 \quad (5.8)$$

CASE 2: $f > f_c$, or $x > 1$. In this case, (5.4) may be written as

$$I_1 = -(2x^2 - 1 + 2x \sqrt{x^2 - 1}) I_2 \quad (5.9)$$

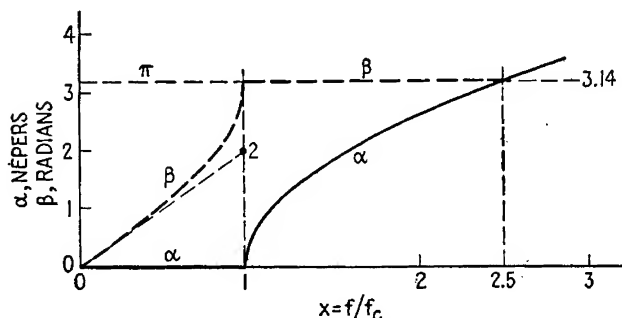


FIG. 5.2.—Variations of the attenuation α and the phase lag β of a one-section low-pass filter terminated by its characteristic impedance.

The quantity in parentheses is now a positive real, increasing from 1 to $+\infty$ as x increases from 1 to ∞ , that is, as f increases from f_c to ∞ . The minus sign indicates that the currents I_1 and I_2 are in phase opposition, or $\beta = \pi$. In this case, Eq. (5.6) may be written as

$$I_1 = -e^\alpha I_2 = e^{\alpha + j\pi} I_2 \quad (5.10)$$

Thus the one general formula (5.6) condenses the two specific ones (5.8) and (5.10). It may also be written

$$I_2 = e^{-\alpha - j\beta} I_1 \quad (5.11)$$

The number α is called the *attenuation* of the filter in *nepers*. In the low-pass filter α is zero for $f < f_c$ and positive for $f > f_c$. The number β is the *phase lag* in radians introduced by the filter. In the low-pass filter, β increases from 0 to $+\pi$ when f increases from 0 to f_c and remains constant and equal to π for $f > f_c$.

Figure 5.2 shows α and β as functions of frequency. The first portion of the β -curve has the form of one-quarter of a sine curve, the tangent at 0 going through the point (1,2). The attenuation α is equal to π nepers for $x = 2.50$.

From (5.2) the voltage ratio is the same as the current ratio. The vector diagram of a one-section filter terminated in Z_c is given in Fig. 5.3.

Below cutoff, Fig. 5.3a, E_1 and I_1 are in phase, Z_c being a pure resistance ($0.866 \sqrt{L/C}$ at $f = f_c/2$). E_2 and I_2 are also in phase, but lagging E_1 and I_1 (for example, $\pi/3$ radians at $f = f_c/2$). E_2 has the same magnitude as E_1 , and I_2 has the same magnitude as I_1 . Power is transmitted from the generator to the load without dissipation in the filter. The frequency band from 0 to f_c is called the *transmission band*.

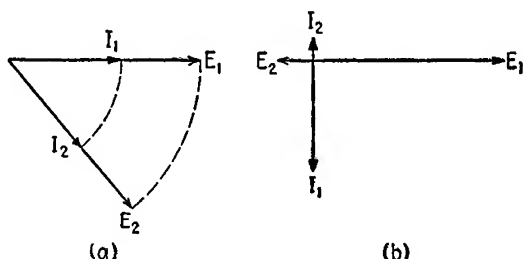


FIG. 5.3.—Vector diagrams of the input and output voltages and currents of a one-section low-pass filter terminated in Z_c : (a) below cutoff, (b) above cutoff.

Above cutoff, Fig. 5.3b, I_1 lags E_1 by $\pi/2$ radians, and I_2 lags E_2 by the same angle, Z_c being a pure reactance. E_2 lags E_1 by π radians since $\beta = \pi$. The ratio of I_2 to I_1 is the same as the ratio of E_2 to E_1 . The frequency band from f_c to infinity is called the *attenuation band*. No power is accepted at the input terminals in the attenuation band since the whole network is purely reactive, the power factor being zero.

In general, any frequency band within which the characteristic impedance is a pure resistance is a band in which the filter accepts power from the generator and transmits all of it to the load, according to the relation

$$\frac{E_1}{I_1} = \frac{E_2}{I_2} = R_c \quad (5.12)$$

Any frequency band within which the characteristic impedance is a pure reactance is a band in which the filter will take no power

from the generator, according to the relation

$$\frac{E_1}{I_1} = \frac{E_2}{I_2} = jX_c \quad (5.13)$$

The definitions and statements made in this section presuppose that the filter is terminated by its characteristic impedance.

6. Low-pass Filter of Several Identical Sections.—The design of a low-pass filter usually involves three factors: the load resistance R , the cutoff frequency f_c , and the sharpness of cutoff as specified by the required attenuation of the filter, say D db, at a frequency a specified percentage above cutoff. If this last condition is not satisfied by the one-section filter discussed in Sec. 5, two or more identical sections may be connected in cascade as explained in Sec. 3. The calculation of the attenuation and phase angle of the ladder-type filter thus obtained is greatly simplified if the filter is assumed to be terminated by its characteristic impedance Z_c instead of by the constant resistance R . In this case each one of the successive filter sections is terminated by Z_c , and

$$E_2 = E_1 e^{-\alpha - i\beta} \quad E_3 = E_2 e^{-\alpha - i\beta} \quad E_4 = E_3 e^{-\alpha - i\beta} \quad \dots \quad (6.1)$$

where E_1 is the input voltage of the first section and E_2, E_3, E_4 are the output voltages of the first, second, third sections. Hence,

$$\begin{aligned} E_3 &= E_1 e^{-2\alpha - 2i\beta} \\ E_4 &= E_1 e^{-3\alpha - 3i\beta} \end{aligned} \quad (6.2)$$

and the attenuation of a filter composed of n identical filter sections is $n\alpha$, and its phase lag is $n\beta$.

Thus the ladder type approximates an ideal filter more and more closely as the number of sections increases, since the attenuation at all frequencies above cutoff becomes greater and greater.

A filter composed of three sections, Fig. 3.2, will contain four coils and three capacitors; and the question arises whether another design, less simple perhaps but using the same number of elements, may not give better results. The composite filters, mentioned in Sec. 19, give one answer to this problem. However, in this chapter only the most fundamental types of filters ("prototypes") will be discussed.

7. Low-pass Filter as Delay Network.—In Chap. IX, Sec. 14, the following relation is obtained:

$$t_d = \frac{\beta}{\omega} \quad (7.1)$$

where t_d is the *time delay* in seconds between two voltage waves of frequency ω , one lagging the other by β radians.

In (5.7) the phase lag β in the transmission band is given by

$$1 - 2x^2 = \cos \beta \quad (7.2)$$

with $x = f/f_c = \omega/\omega_c$. For small values of β ,

$$\cos \beta \doteq 1 - \frac{\beta^2}{2} \doteq 1 - \frac{1}{2}(2x)^2 \quad (7.3)$$

whence

$$\beta \doteq 2x \quad (7.4)$$

and

$$t_d \doteq \frac{2x}{\omega} = \frac{2}{\omega_c} = \sqrt{LC} \quad (7.5)$$

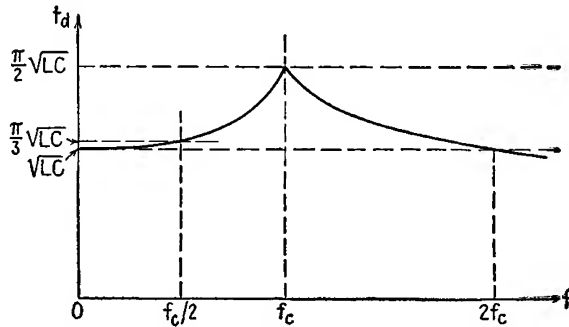


FIG. 7.1.—Variation with frequency of the time delay t_d of a low-pass filter terminated in Z_c .

Figure 7.1 shows that this relation holds with sufficient approximation nearly up to $x = 0.5$, and in this frequency band there is also little attenuation. The low-pass filter is thus a good delay network for periodic signals whose important components lie below $f_c/2$. An n -section filter would have a time delay of $n\sqrt{LC}$ sec in the same band.

8. High-pass Filter.—The simplest high-pass filter consists of two capacitors in series, each of value $2C$, and one inductance L in shunt, Fig. 8.1. For calculations, it is assumed that the filter is terminated by its characteristic impedance Z_c . If

$$\omega_c = \frac{1}{2\sqrt{LC}} \quad \frac{\omega}{\omega_c} = x \quad R = \sqrt{\frac{L}{C}} \quad (8.1)$$

then

$$Z_c = R \sqrt{1 - \frac{1}{x^2}} \quad (8.2)$$

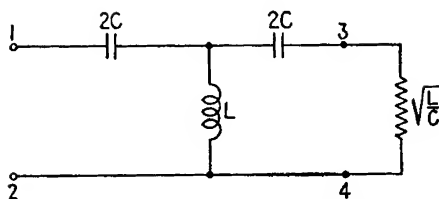


FIG. 8.1.—One-section high-pass filter.

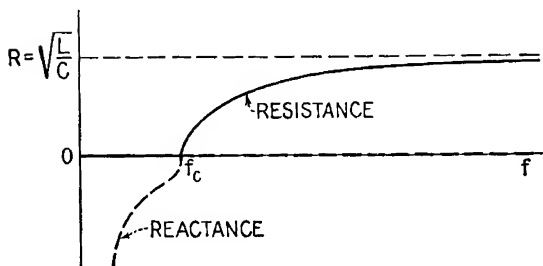


FIG. 8.2.—Variations with frequency of the characteristic impedance of a high-pass filter.

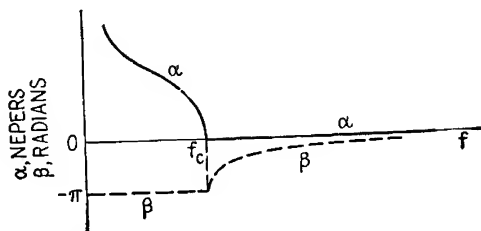
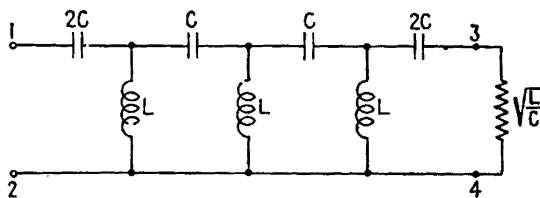
FIG. 8.3.—Variations with frequency of the attenuation α and the phase lag β of a one-section high-pass filter terminated in Z_c .

FIG. 8.4.—A three-section high-pass filter.

Below cutoff, Z_c is a pure negative reactance,¹ decreasing in magnitude from infinity to zero as f increases from zero to f_c . Above cutoff, Z_c is a pure resistance, increasing from zero to R , Fig. 8.2, as f increases from f_c to infinity. The upper frequency band $f > f_c$ is now the *transmission band*, and the lower frequency band

¹ Here Z_c is taken as a negative reactance from the same reasoning that Z_c was taken as a positive reactance in the low-pass filter (see footnote, p. 215).

$f < f_c$ is the *attenuation band*. The attenuation α and phase lag β as functions of frequency are given in Fig. 8.3.

A three-section high-pass filter in ladder-type form is shown in Fig. 8.4. The high-pass filter cannot be used as a delay network.

9. Properties of Two-terminal Nondissipative Networks.—In order to generalize the low-pass and high-pass filters it is necessary

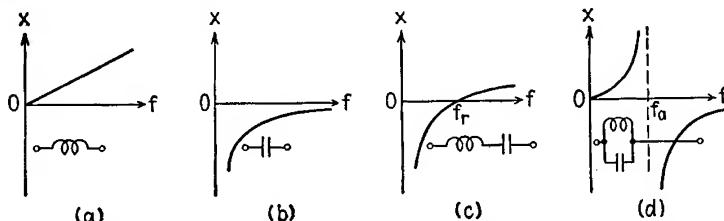


FIG. 9.1.—Variations with frequency of the reactance of four simple networks; f_r and f_a denote resonant and antiresonant frequencies.

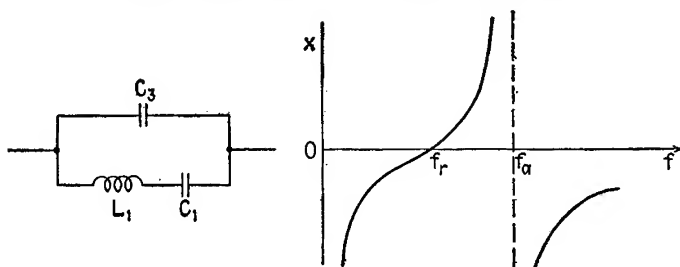


FIG. 9.2.—Variations with frequency of the reactance of the network shown; f_r and f_a denote resonant and antiresonant frequencies.

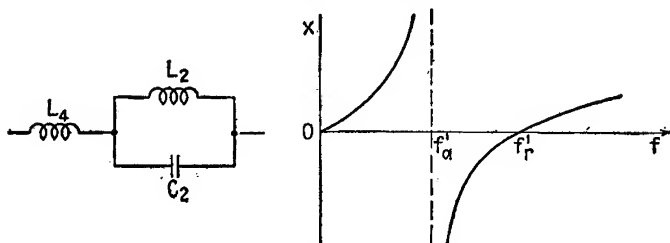


FIG. 9.3.—Variations of the reactance of a network dual to that of Fig. 9.2.

to know certain properties of nondissipative networks. The simplest types of such networks are indicated in Fig. 9.1, which also shows the variations with frequency of their reactance.

In general, there are frequencies for which the reactance is zero, Fig. 9.1c, and other frequencies for which the reactance becomes infinite, Fig. 9.1d. Such frequencies (f_r, f_a) are called *resonant* and *antiresonant* frequencies.

It can be proved that the curve representing the reactance X of a nondissipative network as a function of frequency has *everywhere a positive slope* (Campbell-Foster theorem). As examples, Figs. 9.2 and 9.3 show the variations of reactance for two reactive networks of three elements each (these networks occur in the analysis of quartz and magnetostriction oscillators, respectively).

10. Inverse Two-terminal Networks.—In the low-pass filter of Fig. 2.2, the series branch is inductive and the shunt branch is capacitive at all frequencies. In the high-pass filter of Fig. 8.1, the series branch is always capacitive, the shunt branch always inductive. In general, let $Z_1 = jX_1$ and $Z_2 = jX_2$ be the impedances of the series and shunt networks that form the arms of the filter; in the type of filters to be considered now, X_1 and X_2 will be of opposite signs at any given frequency. Two reactive networks that satisfy this condition are called *dual* or *inverse networks*. (The expressions *arm*, *branch*, *two-terminal network* are equivalent.) The “critical frequencies” at which their reactances become zero or infinite are the same, one of the networks being resonant ($X_1 = 0$) at every critical frequency for which the other network is antiresonant ($X_2 = \infty$), and vice versa.

It is a remarkable fact that, if X_1 and X_2 , Figs. 9.1c and d, are of opposite sign at each and every frequency, then the product X_1X_2 is independent of frequency. For inverse networks Z_1 and Z_2 ,

$$Z_1Z_2 = k^2 \quad \text{or} \quad X_1X_2 = -k^2 \quad (10.1)$$

By comparing the circuit diagrams of two inverse networks, Figs. 9.2 and 9.3, for example, it is seen that for every L in one network there is a C in the other, and vice versa. Moreover, two elements that are in series in one network are in parallel in its inverse, and vice versa. It can be shown that these topological conditions are necessary but not sufficient in the general case in order that two reactive networks composed of series and parallel branches be dual. It remains to choose the numerical values of the L 's and C 's so that every resonant frequency of one network is equal to an antiresonant frequency of the other network, and vice versa.

The numerical values of the L 's and C 's must be such that the ratio of any L in one network to the corresponding C in the other network is the same for all element pairs. For example, for Figs. 9.2 and 9.3, these conditions are

$$\frac{L_1}{C_2} = \frac{L_2}{C_1} = \frac{L_4}{C_3} = k^2 \quad (10.2)$$

Then the networks are inverse, and the common value of the ratios (10.2) is the constant k^2 in (10.1). Note that k has the dimensions of resistance.

11. Constant- k Filters.—Constant- k filters are built with inverse reactive networks Z_1 and Z_2 as their series and shunt arms, Figs. 11.1 and 11.2. The low- and high-pass filters of Secs. 3 and 8 are the simplest examples of constant- k filters. In general, the product $Z_1 Z_2 = k^2$ is taken equal to R^2 , where R is the value

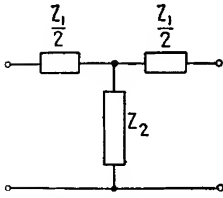


FIG. 11.1.—A one-section constant- k filter.

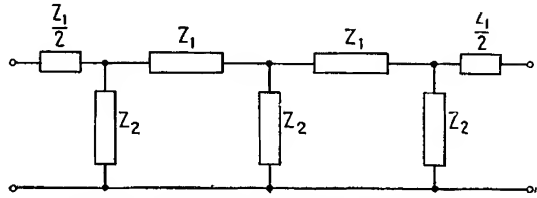


FIG. 11.2.—A three-section constant- k filter.

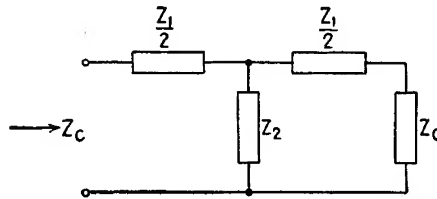


FIG. 11.3.—A one-section constant- k filter terminated in its characteristic impedance.

of the resistive load. This simple value gives a good frequency discrimination.

The characteristic impedance Z_c is obtained as in Sec. 4 by equating the input impedance of the filter terminated in Z_c to Z_c itself, Fig. 11.3,

$$Z_{in} = Z_c = \frac{Z_1}{2} + \frac{\left(\frac{Z_1}{2} + Z_c\right) Z_2}{\frac{Z_1}{2} + Z_c + Z_2} \quad (11.1)$$

Solving for Z_c , using the relation $Z_1 Z_2 = R^2$,

$$Z_c = R \sqrt{1 + \frac{Z_1}{4Z_2}} \quad (11.2)$$

The ratio $Z_1/4Z_2$, which is a function of frequency, plays a fundamental role in constant- k filters. The reactances X_1 and X_2

are always of opposite sign; and if u is written for $Z_1/4Z_2$,

$$u = \frac{Z_1}{4Z_2} = \frac{X_1}{4X_2} = -\left(\frac{X_1}{2R}\right)^2 \quad (11.3)$$

and

$$Z_c = R \sqrt{1+u} \quad (11.4)$$

The parameter u is always a negative real quantity for a constant k filter and is a function of frequency. In the simple case of a low-pass filter,

$$u = -\left(\frac{\omega L}{2\sqrt{L/C}}\right)^2 = -\left(\frac{\omega}{\omega_c}\right)^2 = -x^2 \quad (11.5)$$

and in a high-pass filter

$$u = -\left(\frac{1}{2\omega C \sqrt{L/C}}\right)^2 = -\left(\frac{\omega_c}{\omega}\right)^2 = -\frac{1}{x^2} \quad (11.6)$$

12. Attenuation Constant and Phase Lag of Constant- k Filters.

The next step is to calculate the attenuation constant α (nepers) and the phase lag β (radians) of one section of the constant- k filter, terminated in its characteristic impedance. From (5.1) and (5.6),

$$\frac{E_1}{E_2} = \frac{I_1}{I_2} = e^{\alpha+j\beta} \quad (12.1)$$

and after some calculation this ratio is found to be

$$e^{\alpha+j\beta} = 2u + 1 + \sqrt{4u(1+u)} \quad (12.2)$$

The filter behaves in two fundamentally different ways according to whether the magnitude of u is less or greater than unity.

Transmission Bands, $-1 < u < 0$.—The quantity under the radical in (12.2) is negative, and the entire complex quantity (12.2) has a magnitude equal to unity, so that

$$\alpha = 0 \quad (12.3)$$

Then

$$e^{j\beta} = (2u + 1) + j\sqrt{-4u(1+u)} \quad (12.4)$$

with

$$\cos \beta = 2u + 1 \quad (12.5)$$

$$\sin \beta = \sqrt{-4u(1+u)} \quad (12.6)$$

For frequencies such that $-1 < u < 0$, there is thus no attenuation. Such frequency bands are the *transmission bands*.

Attenuation Bands, $u < -1$.—The quantity (12.2) is a negative real number, so that $\epsilon^{i\beta} = -1$ and $\beta = \pm\pi$, or more generally $\beta = (2n + 1)\pi$, and

$$\epsilon^\alpha = (-2u - 1) + \sqrt{4u(1 + u)} \quad (12.7)$$

The calculation of α from (12.7) is much simplified by introducing a function of α , called the hyperbolic cosine, abbreviated $\cosh \alpha$, and defined by

$$\cosh \alpha = \frac{\epsilon^\alpha + \epsilon^{-\alpha}}{2} \quad (12.8)$$

From (12.7),

$$\epsilon^{-\alpha} = (-2u - 1) - \sqrt{4u(1 + u)} \quad (12.9)$$

whence

$$\cosh \alpha = -1 - 2u \quad (12.10)$$

The value of α when u is known is obtained from a table of hyperbolic cosines, commonly found in mathematical and communication handbooks.

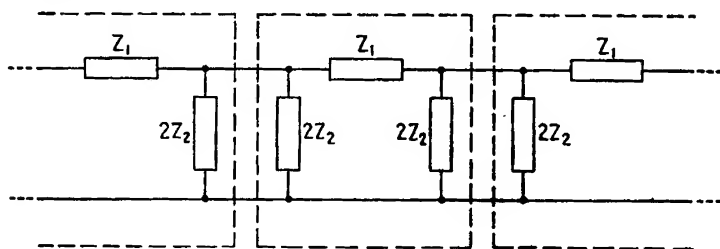


FIG. 12.1.—Derivation of the critical condition for a constant- k filter.

Frequency bands for which $u = Z_1/4Z_2 < -1$ are *attenuation bands*. Thus $u = -1$ is the critical value of the parameter u , separating the transmission bands from the attenuation bands. At the cutoff frequencies, $u = Z_1/4Z_2 = -1$, and

$$Z_1 + 4Z_2 = 0 \quad (12.11)$$

The following remark suggests a physical reason for this condition. Consider a ladder-type filter of a number of sections, Fig. 12.1. Each shunt network Z_2 can be replaced by two networks of impedance $2Z_2$ in parallel. The filter will thus consist of a number of identical “cells” (each enclosed in a box in Fig. 12.1). Within each cell there is a “loop” formed by the network Z_1 in series (around the loop) with a network of impedance $4Z_2$. A critical condition will obtain at any frequency for which each cell would resonate if it were an isolated loop, giving condition (12.11).

13. Band-pass Filter.—The preceding general discussion will now be applied to a constant- k filter with one transmission band. The diagram of such a filter contains a series-resonant circuit in series, and a parallel-antiresonant circuit in shunt, Fig. 13.1. As described in Secs. 10 and 11,

$$\frac{L_1}{C_2} = \frac{L_2}{C_1} = k^2 = R^2 \quad (13.1)$$

The variations of the series and shunt reactances X_1 and X_2 are given by Fig. 13.2, where $4X_2$ has been plotted instead of X_2 . The

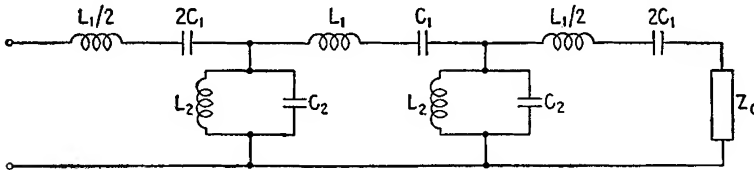


FIG. 13.1.—A two-section constant- k band-pass filter.

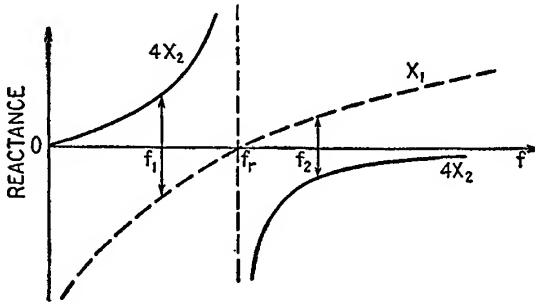


FIG. 13.2.—Variations with frequency of X_1 and $4X_2$.

common resonant frequency of the series and shunt networks is

$$f_r = \frac{\omega_r}{2\pi} \quad \text{and} \quad \omega_r = \frac{1}{\sqrt{L_1 C_1}} = \frac{1}{\sqrt{L_2 C_2}} \quad (13.2)$$

The parameter u as defined in (11.4) is

$$u = \frac{X_1}{4X_2} \quad (13.3)$$

Therefore, calling f_1 the frequency at which $4X_2 = -X_1 > 0$ and f_2 the frequency at which $X_1 = -4X_2 > 0$, as in Fig. 13.2, u has a magnitude greater than unity for $f < f_1$ and $f > f_2$, and a magnitude less than unity for $f_1 < f < f_2$. Therefore, there is *one transmission band* between f_1 and f_2 with one attenuation band on each side.

For any given frequency, numerical values of α and β are easily obtained by using (12.5) and a table of cosines for β , and (12.10) and a table of hyperbolic cosines for α . The variations of these two quantities are sketched in Fig. 13.3. The attenuation α is zero between f_1 and f_2 and rises on either side vertically at first. The output quantities E_2, I_2 lag the input quantities E_1, I_1 at frequencies greater than the resonant frequency f_r (as in a low-pass filter) and lead E_1, I_1 at frequencies less than f_r (as in a high-pass filter). It can be shown that

$$f_r = \sqrt{f_1 f_2} \quad (13.4)$$

The practical problem is to design a band-pass filter, to pass frequencies between f_1 and f_2 , working into a constant resistance R . The attenuation curve α for such a filter will not show right angles

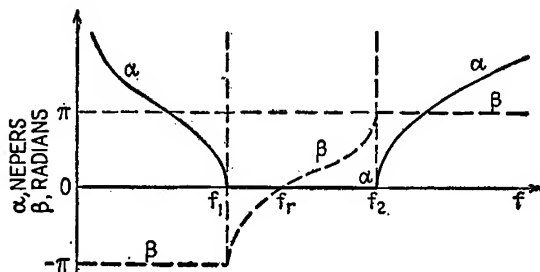


FIG. 13.3.—Variations with frequency of the attenuation α and the phase lag β of a constant- k band-pass filter terminated in Z_c .

as in Fig. 13.3; the corners will be rounded off, just as the right angle of the α -curve of Fig. 5.2 is shown rounded off in Fig. 2.3. The load resistance R usually is taken equal to the square root of the ratios in (10.2) or (13.1). Upon calculating the values f_1 and f_2 for which $X_1 + 4X_2 = 0$, Fig. 13.2, the values of the design constants are

$$\left. \begin{aligned} L_1 &= \frac{R}{\pi(f_2 - f_1)} & L_2 &= \frac{R(f_2 - f_1)}{4\pi f_1 f_2} \\ C_1 &= \frac{f_2 - f_1}{4\pi R f_1 f_2} & C_2 &= \frac{1}{\pi R(f_2 - f_1)} \end{aligned} \right] \quad (13.5)$$

These formulas are equivalent to the following procedure:

1. Design a low-pass filter with a cutoff $f_c = f_2 - f_1$ and a load R . Its elements will be L_1 and C_2 .
2. Design a high-pass filter with a cutoff $f_c = f_1 f_2 / (f_2 - f_1)$ and a load R . Its elements will be L_2 and C_1 .

14. Band-stop Filters.—It is sometimes desired to design a filter that will stop the frequencies higher than f_1 and lower than f_2 , that is, a *band-stop filter*. The circuit for such a filter is easily obtained from the application of the general principle that, if the series and shunt networks of a filter are interchanged, the new filter will have the same number of bands as the original filter, a transmission band taking the place of an attenuation band, and vice versa. For example, a filter with an antiresonant circuit in series and a resonant circuit in shunt is a band-stop filter. However, the cutoff frequencies will not be the same as those of the original filter, since they occur when $Z_1 + 4Z_2 = 1$. Therefore, for the same cutoff frequencies, the L 's and C 's of the band-stop filter must be recalculated.

15. II-section Filters.—A one-section filter may be connected in the form of a Π , instead of the T used previously in this chapter. The series impedance is labeled Z_1 , the shunt impedances $2Z_2$, Fig. 15.1. A filter composed of several sections will have Z_1 and Z_2 for series and shunt impedances, respectively, except at the ends, Fig. 15.2, where the shunt impedances will be $2Z_2$.

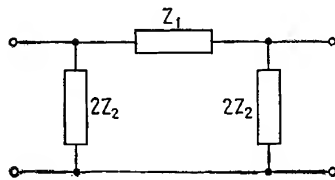


FIG. 15.1.—Symmetrical II-section filter.

It can be shown that the values of α and β for the Π -section filter are the same at all frequencies as for the T -section filter having the same Z_1 and Z_2 , and the transmission bands will be the same. The characteristic impedance is different and is given by

$$Z_c = \frac{R}{\sqrt{1 + u}} \quad (15.1)$$

However, the values of Z_c for the T - and the Π -filter are the same and equal to R for the frequencies corresponding to $u = 0$ (which

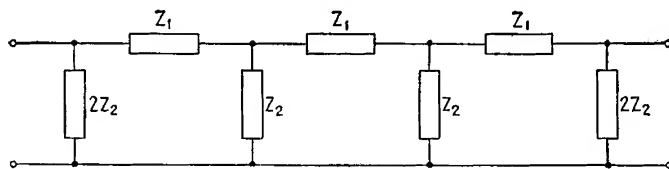


FIG. 15.2.—Three-section Π filter.

are zero for low-pass, infinity for high-pass, f_r for band-pass); and therefore the same design formulas hold for the Π - and the T -section when the filter is terminated by a constant load R .

16. Lattice Structures.—Lattice-type filters, Fig. 16.1, are sometimes used. The characteristic impedance Z_c is

$$Z_c = \sqrt{Z_1 Z_2} \quad (16.1)$$

the phase lag β in the transmission bands is given by

$$\tan \beta = \frac{2\sqrt{-u}}{1+u} \quad \text{or} \quad \tan \frac{\beta}{2} = \sqrt{-u} \quad (16.2)$$

and the attenuation α in the attenuation bands by

$$\tanh \alpha = \frac{2\sqrt{-u}}{1+u} \quad (16.3)$$

whence it is easy to obtain the frequency characteristics for a lattice filter when X_1 and X_2 are given.

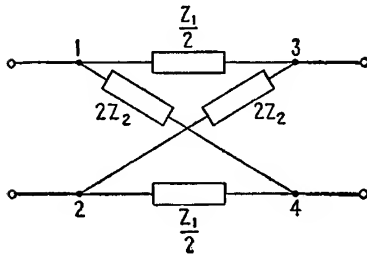


FIG. 16.1.—A one-section lattice filter.

A remarkable case is that of the constant- k lattice filter, in which Z_1 and Z_2 are inverse networks such that $Z_1 Z_2 = R^2$. Then $Z_c = R$ at all frequencies; hence there is transmission at all frequencies, Sec. 5. The phase lag is given by (16.2). Such *all-pass* networks have no other effect than to introduce a phase shift that is a function of frequency, and this may be of interest, in particular for a delay network.

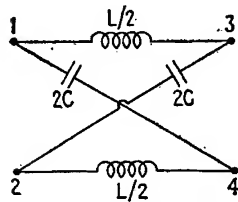


FIG. 16.2.—A lattice-shaped delay network.

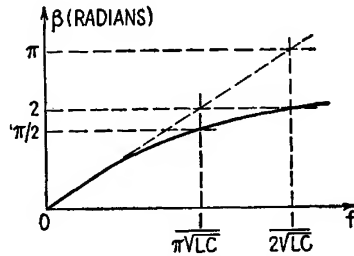


FIG. 16.3.—Phase-lag characteristic of the delay network of Fig. 16.2.

As an example, the network of Fig. 16.2 has zero attenuation at all frequencies and introduces a phase lag β given by

$$\tan \frac{\beta}{2} = \frac{\omega \sqrt{LC}}{2} \quad (16.4)$$

and shown in Fig. 16.3. The time delay $t_d = \beta/\omega$ is approximately constant at low frequencies and equal to \sqrt{LC} . This value is the same as for a one-section low-pass ladder-type filter, Sec. 7, but the lattice network will introduce less distortion.

17. Effect of Losses.—As mentioned in Sec. 1, all the filter elements have been assumed dissipationless in the previous discussions. The effect of losses in either inductors or capacitors is to round off the corners of the α - and β -characteristics at the cutoff frequencies. This effect is usually small. Where it must be taken into account, the filter is recalculated or its characteristics measured after its design has been first determined on a resistanceless basis. For the constant- k filter, correction curves have been calculated.¹

18. Other Types of Filter Circuits.—The arms of a lattice-filter section, Fig. 16.1, are actually the four arms of a bridge, of which

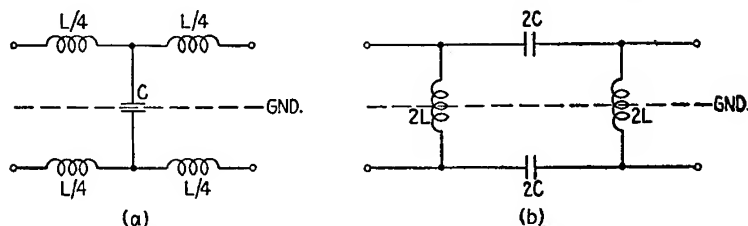


FIG. 18.1.—(a) Low-pass filter, H type, $\omega_c = 2/\sqrt{LC}$; (b) high-pass filter, square type, $\omega_c = 1/(2\sqrt{LC})$.

1-2 and 3-4 are opposite corners. The condition of balance of the bridge is $Z_1 = 4Z_2$, or $u = 1$. For a constant- k all-pass lattice filter, u is always negative, from (11.3), and the bridge is always unbalanced.

All the filters discussed in Secs. 1 to 15 are four-terminal networks only in appearance, since an equipotential connection was assumed between terminals 3 and 4. In certain cases, particularly with transmission lines, it is necessary to rearrange the parts to secure symmetry with regard to a ground plane, actual or assumed. The result is a *balanced filter*, of the H or square type according to whether the corresponding unbalanced filter is of the T or Π type. Figures 18.1a, b show balanced filters of the low-pass H type and high-pass square type. Symbols are chosen so that the formulas for ω_c and for the resistive load R are the same as for the unbalanced type. Lattice filters are always of the balanced type.

¹ T. E. SHEA, "Transmission Networks and Wave Filters," Sec. 40, D. Van Nostrand Company, Inc., New York, 1929.

Frequency discrimination also can be obtained by using resonant circuits, which are discussed in other chapters. For example, a narrow band-pass or band-stop filter can be made with one resonant or antiresonant circuit. A band-pass filter can be made with two resonant coupled circuits (see Chap. VII).

19. Low-pass Composite Filter.—A one-section resistanceless filter gives poor frequency discrimination for two reasons: (1) In

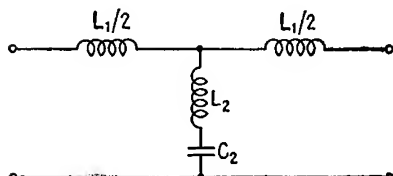


FIG. 19.1.—Example of *m*-derived filter section (No. 2 of Fig. 19.3).

the attenuation band, the attenuation does not increase very fast, *i.e.*, the cutoff is not sharp enough. (2) In the transmission band, the characteristic impedance is a very poor match for the usual resistive load. Building a filter from two or three identical sections

in cascade remedies the first defect to a certain extent but requires numerous coils and capacitors. A better result for a given number of elements is obtained from a *composite filter*, using *m*-derived *sections* of a more elaborate design than the *prototypes* that have been described. The theory of composite filters will not be given here; only a specific composite low-pass filter will be designed as an example.

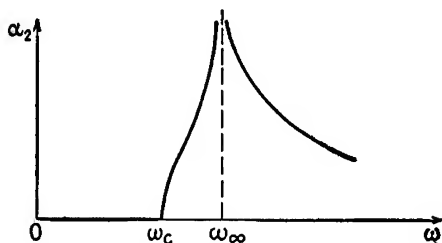


FIG. 19.2.—Attenuation constant of *m*-derived filter section.

Consider first the one-section filter of Fig. 19.1. This is a low-pass filter with a cutoff ω_c occurring when

$$Z_1 + 4Z_2 = 0 \quad (19.1)$$

Hence,

$$\omega_c = \frac{2}{\sqrt{(L_1 + 4L_2)C_2}} \quad (19.2)$$

The attenuation α_2 of this filter is zero below cutoff, Fig. 19.2. In the attenuation band it becomes infinite at the frequency

$$\omega_{\infty} = \frac{1}{\sqrt{L_2 C_2}} \quad (19.3)$$

In order to use this filter section in cascade with a prototype low-pass filter, it is desirable that their characteristic impedances be the same at all frequencies. If L and C are the elements of the prototype, it is found that this condition is fulfilled when

$$\left. \begin{aligned} L_1 &= mL \\ L_2 &= \frac{1-m^2}{4m} L \\ C_2 &= mC \end{aligned} \right\} \quad (19.4)$$

With these values,

$$\omega_c = \frac{2}{\sqrt{LC}} \quad \omega_{\infty} = \frac{\omega_c}{\sqrt{1-m^2}} \quad m = \sqrt{1 - \left(\frac{\omega_c}{\omega_{\infty}}\right)^2} \quad (19.5)$$

Thus, upon choosing for ω_{∞} any value of frequency above cutoff, a numeric m results that lies between 0 and 1, and the filter of Fig.

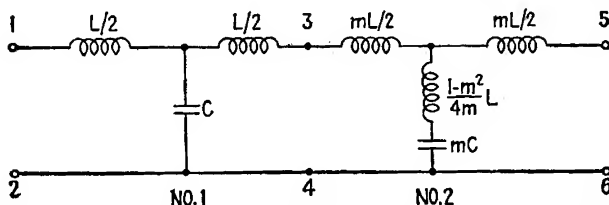


FIG. 19.3.—Two-section filter, with one prototype and one m -derived section.

19.3 composed of the prototype (No. 1) and the m -derived section (No. 2) in cascade has ω_c for its cutoff and its attenuation is $\alpha = \alpha_1 + \alpha_2$. Since α_2 is infinite at the angular frequency ω_{∞} , the total α -curve, Fig. 19.4, will be much steeper in the neighborhood of the cutoff than the $(2\alpha_1)$ -curve for a filter composed of two identical No. 1 prototypes.

It still remains to correct the image impedance of the network so that it is more nearly constant below cutoff. A rather intricate theory shows that the best result is obtained from the following simple rule:

1. Design an m -derived filter from (19.4), using for m the value $m = 0.6$.
2. Construct the Π -type section having the same series and shunt impedances Z_1, Z_2 as in (1), Fig. 19.5.

3. Place the left half of the Π at the left of the terminals 1-2 in Fig. 19.3 and the right half at the right of terminals 5-6.

As an example, a composite low-pass filter cutting off at 1,000 cps, working into a load of 500 ohms, consisting of the prototype section, one m -derived section ($f_\infty = 1,200$, $m = 0.553$), and two half- Π sections ($m = 0.6$), will appear as in Fig. 19.6, after reduction to the smallest number of elements.

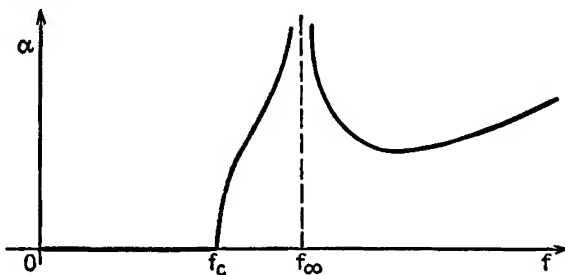


FIG. 19.4.—Over-all attenuation characteristic of one prototype section and one m -derived section in cascade, as in Fig. 19.3.

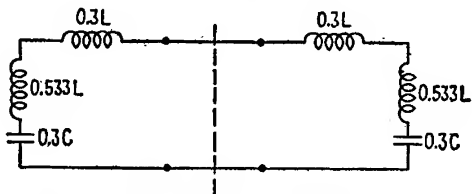


FIG. 19.5.—Example of m -derived Π section.

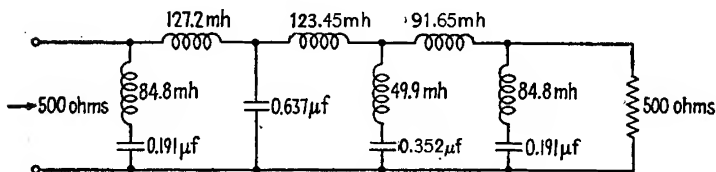


FIG. 19.6.—Composite filter.

The “body” of the filter, consisting in Fig. 19.3 of the prototype and one m -derived section, in general may contain two or more m sections calculated, for example, for $f_\infty/f_c = 1.25$, 1.6, 2 and thus may have a very high attenuation in a broad band above cutoff, the two “ends” of the composite filter remaining in any case as in Fig. 19.5.

20. Applications.—Some important applications of frequency discrimination in radio are for filtration of ripples in power supplies,

limitation of band width in i-f amplifiers, broadening of band width in certain types of video amplifiers, and filtering of control signals mixed with the program in broadcasting.

In communication, carrier-current telephony and acoustical filtering are the major fields of application.

Supplementary Reading

- K. S. JOHNSON: "Transmission Circuits for Telephone Communication," 6th ed., D. Van Nostrand Company, Inc., New York, 1937.
- T. E. SHEA: "Transmission Networks and Wave Filters," D. Van Nostrand Company, Inc., New York, 1929.
- W. L. EVERITT: "Communication Engineering," 2d ed., Chap. VI, McGraw-Hill Book Company, Inc., New York, 1937.
- E. A. GUILLEMIN: "Communication Networks," Vol. II, John Wiley & Sons, Inc., New York, 1935.
- H. PENDER and K. McILWAIN: "Electrical Engineers' Handbook," 3d ed., Vol. V, John Wiley & Sons, Inc., New York, 1936.
- L. H. WARE and H. R. REED: "Communication Circuits," Chaps. IX, X, John Wiley & Sons, Inc., New York, 1942, 2d ed., 1944.
- F. E. TERMAN: "Radio Engineers' Handbook," Sec. 3, par. 28, 29, McGraw-Hill Book Company, Inc., New York, 1943.
- L. BRILLOUIN: "Wave Propagation in Periodic Structures," McGraw-Hill Book Company, Inc., New York, 1946.

CHAPTER IX

FOURIER ANALYSIS

1. Periodic Function.—A periodic function (of time, for example) is defined by this property, that the value of the function at time $t + T$ is the same as the value of the function at any time t . The constant T is an interval of time called the *period* and is the shortest interval for which the above property holds. The graph of a periodic function is an arc AB , of whatever shape, Fig. 1.1, which repeats itself indefinitely. When the period is an interval of time T , its inverse $1/T = f$ is called the *frequency*. The independent variable may be distance instead of time. Then the period is a length λ called the *wavelength*. The succession of values of time or of distance within a period is called a *cycle*.

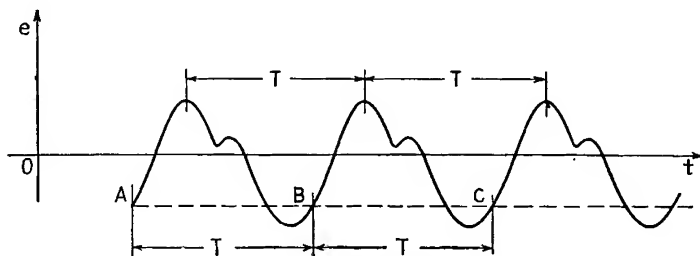


FIG. 1.1.—Periodic function of time.

Examples of periodic functions are $\sin x$, $\cos x$. Their period is 2π radians. An example of a periodic function of time, of period T , is $\sin 2\pi t/T$. The quantity $\omega = 2\pi/T$ is the angular velocity; and, in electrical problems, $x = \omega t$ is called the electrical angle.

2. Object of Fourier Analysis.—If a harmonic voltage such as $\sin \omega t$, $\cos \omega t$, $\cos (\omega t - \phi)$, is applied between two terminals of a linear network, all potential differences and currents in the network are harmonic and of the same angular frequency ω . In the study of a-c theory and networks it is shown how to calculate the amplitudes and phases of these potential differences and currents when the applied emf is known. However, in communication engineering, periodic oscillations occur that are not harmonic. The microphone current when a steady sound is sung or spoken,

the output of a detector on which a harmonic voltage is impressed, the scanning voltage on a cathode-ray tube are important examples of periodic nonharmonic oscillations. Fourier analysis shows that a nonharmonic wave, Fig. 1.1, consists of harmonic components, of which the component of lowest frequency, or fundamental, has the same period T as the nonharmonic wave. The other components have periods of $T/2$, $T/3$, etc. A nonharmonic wave consists of a number of harmonic waves of frequencies $f = 1/T$, $2f$, $3f$, The wave of frequency f is called the fundamental; the waves of frequencies $2f$, $3f$, . . . are called the second harmonic, third harmonic,

Assume that a nonharmonic voltage is applied to a certain network and that the output voltage is desired. Ordinary a-c theory may be used to obtain the harmonic output voltage due to any one of the harmonic components of the input voltage. If the network is linear, the principle of superposition holds and the actual output is the sum of all the harmonic output components. The building up of a periodic function from its harmonic components is called Fourier synthesis.

A network problem involving periodic nonharmonic voltages and currents may be solved in three stages: (1) Fourier analysis of the input voltage or current; (2) calculations on the network, according to a-c theory, for each component frequency; (3) Fourier synthesis of the output.

3. Even and Odd Components (C and S Components).—The periodic nonharmonic voltage $e(t)$ to be analyzed is supposed to be

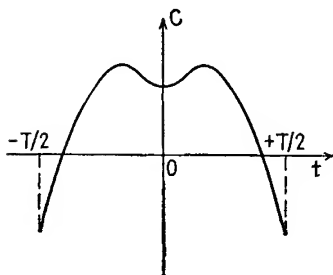


FIG. 3.1.—Example of an even function, or C function.

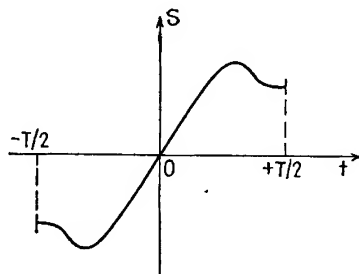


FIG. 3.2.—Example of an odd function, or S function.

known for a whole period of time T , from $t = -T/2$ to $t = +T/2$, let us say, either through some mathematical formula or from some measured or recorded curve. The first step will be to analyze $e(t)$ into the sum of an odd and an even component.

A symmetrical function is one of which the graph is symmetrical with regard to the vertical axis, Fig. 3.1. Examples of symmetrical functions are t^2 , t^4 , t^{-2} , . . . and $\cos \omega t$, $\cos 2\omega t$, $\cos \omega t/2$,

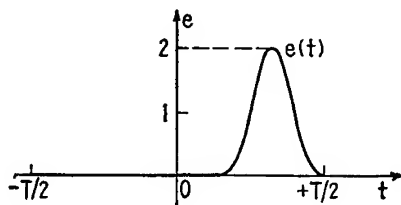


FIG. 3.3.—An arbitrary function of time, $e(t)$, between $-T/2$ and $+T/2$.

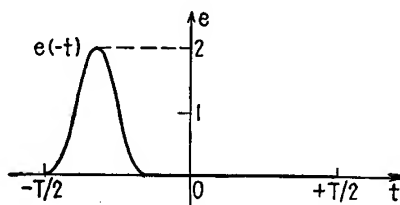


FIG. 3.4.—The function $e(-t)$.

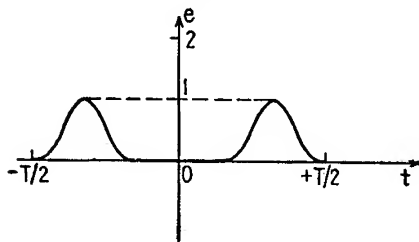


FIG. 3.5.—The even or C component of $e(t)$.

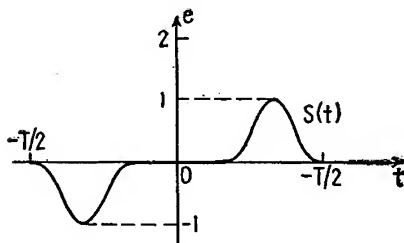


FIG. 3.6.—The odd or S component of $e(t)$.

Functions that are symmetrical with respect to the vertical axis are commonly called *even* functions or C functions.

An antisymmetrical function is one of which the graph is symmetrical with regard to the origin O , Fig. 3.2. Examples of anti-

symmetrical functions are t , t^3 , t^{-1} , . . . and $\sin \omega t$, $\sin 2\omega t$, $\sin \omega t/2$, Functions that are symmetrical with respect to the origin are commonly called *odd* functions or *S* functions.

The required analysis of the periodic voltage $e(t)$ results from

$$\begin{aligned} e(t) &= \frac{e(t) + e(-t)}{2} + \frac{e(t) - e(-t)}{2} \\ &= C(t) + S(t) \end{aligned} \quad (3.1)$$

which shows that any function $e(t)$ is the sum of a C component and an S component. An example of such an analysis is shown in Figs. 3.3 to 3.6.

4. Analysis of an Even Function $C(t)$.—The next step is to analyze an even function $C(t)$ into its frequency components. Assume that the analysis is possible, or

$$C(t) = B_0 + B_1 \cos \omega t + B_2 \cos 2\omega t + \cdots + B_n \cos n\omega t + \cdots \quad (4.1)$$

The constant term B_0 is often called the d-c component, $B_1 \cos \omega t$ is the fundamental component or first harmonic, $B_2 \cos 2\omega t$ is the second harmonic, etc.

The fundamental angular velocity ω is $\omega = 2\pi/T$, and the frequency $f = \omega/2\pi$ is the fundamental frequency.

To obtain B_0 , consider, on the one hand, the area in Fig. 4.1 limited by the axis Ot , the graph of $C(t)$, and the verticals $t = -T/2$, $t = +T/2$; on the other hand, the areas similarly

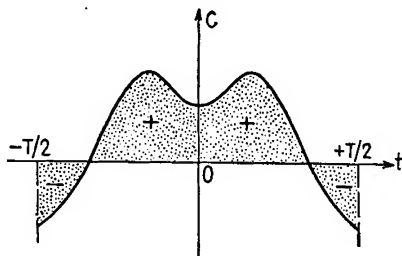


FIG. 4.1.—Area under a C function between $-T/2$ and $+T/2$.

limited by the graphs of the successive terms on the right-hand side of (4.1). These areas are taken as positive when they lie above the Ot axis, negative when they lie below it. The algebraic value of the area relative to $B_1 \cos \omega t$ is zero, and this is true for all the following components in (4.1). Therefore, the area limited by $C(t)$ from $t = -T/2$ to $t = +T/2$ is equal to T times B_0 . Consequently, $B_0 = \text{average value of } C(t) \text{ over one period or cycle or over any whole number of cycles, Fig. 4.2.}$

To obtain B_1 , multiply each side of (4.1) by $\cos \omega t$ and consider the areas limited by the graphs of the different terms. Every term on the right-hand side gives zero area except the term $B_1 \cos^2 \omega t$. The average value of $\cos^2 \omega t$ over one period is $\frac{1}{2}$, Fig. 4.3.

The area under the curve $B_1 \cos^2 \omega t$ for one period is therefore $T B_1/2$. Hence the area limited by $C(t) \cos \omega t$ from $t = -T/2$ to $t = +T/2$ is equal to $T B_1/2$, and $B_1 =$ twice the average value of $C(t) \cos \omega t$ over one period or cycle. Similarly, for $n = 2, 3, \dots, n$, $B_n =$ twice the average value of $C(t) \cos n\omega t$ over one period or cycle. Note that the expression for the d-c component B_0 differs from that for B_1, B_2, \dots, B_n .

In theoretical problems a general expression for B_n should be obtained, if possible. In practice, B_0, B_1, B_2, \dots may be evaluated graphically or experimentally; when the coefficients become smaller than the probable experimental error, they are neglected.

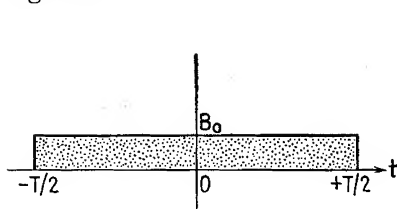


FIG. 4.2.—Average value of $C(t)$ between $-T/2$ and $+T/2$.

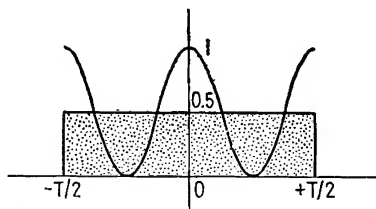


FIG. 4.3.—The $\cos^2 \omega t$ and its average between $-T/2$ and $+T/2$.

5. Odd Function $S(t)$.—The analysis of an odd function $S(t)$ is similar and gives

$$S(t) = A_1 \sin \omega t + A_2 \sin 2\omega t + \dots + A_n \sin n\omega t + \dots \quad (5.1)$$

There can be no constant term, and the formula for all the A 's is $A_n =$ twice the average value of $S(t) \sin n\omega t$ over one period or cycle.

6. Complete Fourier Analysis.—The complete Fourier analysis of a periodic function that is neither even nor odd is, from (3.1),

$$\begin{aligned} e(t) &= C(t) + S(t) \\ &= B_0 + B_1 \cos \omega t + B_2 \cos 2\omega t + \dots \\ &\quad + A_1 \sin \omega t + A_2 \sin 2\omega t + \dots \end{aligned} \quad (6.1)$$

The coefficients B_0, B_n, A_n are given in Secs. 4 and 5. It is not necessary to analyze $e(t)$ into C and S components to obtain B_0, B_n, A_n , because the average value of the odd functions $C(t) \sin n\omega t$ and $S(t) \cos n\omega t$ over one period or cycle is zero. Therefore, $S(t)$ may be added to $C(t)$ or $C(t)$ to $S(t)$ without changing the results of Secs. 4 and 5, and

$$B_0 = \text{average of } e(t) \text{ over one period or cycle} \quad (6.2)$$

$$B_n = \text{twice the average of } e(t) \cos n\omega t \text{ over 1 period or cycle} \quad (6.3)$$

$$A_n = \text{twice the average of } e(t) \sin n\omega t \text{ over 1 period or cycle} \quad (6.4)$$

In integral notation, these formulas are written

$$B_0 = \frac{1}{T} \int_{-T/2}^{+T/2} e(t) dt \quad (6.5)$$

$$B_n = \frac{2}{T} \int_{-T/2}^{+T/2} e(t) \cos n\omega t dt \quad (6.6)$$

$$A_n = \frac{2}{T} \int_{-T/2}^{+T/2} e(t) \sin n\omega t dt \quad (6.7)$$

$B_0/2$ is sometimes written for the d-c component, and then all the B 's have the same form.

7. Amplitude and Phase of Each Harmonic Component.—It is often convenient to combine two harmonic components (sin and cos) of the same frequency into a single component. Let $A \sin \omega t$ and $B \cos \omega t$ be the components to be added. Let A and B be the position at the time $t = 0$ of two rotating vectors, Fig. 7.1. If these vectors rotate counterclockwise with the angular velocity ω , their projections on the vertical axis at the time t will be $A \sin \omega t$ and $B \cos \omega t$. The sum of the two harmonic components will be represented by the resultant of the two vectors A and B , a vector of magnitude $M = \sqrt{A^2 + B^2}$, lying ψ radians ahead of the A vector, with $\tan \psi = B/A$.

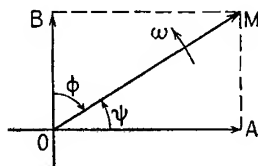


FIG. 7.1.—Rotating vectors representing $A \sin \omega t$ and $B \cos \omega t$.

The same result is obtained algebraically by writing

$$A \sin \omega t + B \cos \omega t = \sqrt{A^2 + B^2} \left(\frac{A}{\sqrt{A^2 + B^2}} \sin \omega t + \frac{B}{\sqrt{A^2 + B^2}} \cos \omega t \right)$$

Putting

$$M = \sqrt{A^2 + B^2} \quad (7.1)$$

$$\cos \psi = \frac{A}{\sqrt{A^2 + B^2}} \quad \sin \psi = \frac{B}{\sqrt{A^2 + B^2}} \quad \tan \psi = \frac{B}{A} \quad (7.2)$$

one has

$$A \sin \omega t + B \cos \omega t = M \sin (\omega t + \psi) \quad (7.3)$$

Instead of the two amplitudes A and B , two new parameters M and ψ appear, an amplitude and a phase angle, related to A and B by (7.1) and (7.2).

The Fourier development (6.1) takes the form

$$e(t) = B_0 + M_1 \sin(\omega t + \psi_1) + M_2 \sin(2\omega t + \psi_2) + \cdots \quad (7.4)$$

It can equally well be written

$$e(t) = B_0 + M_1 \cos(\omega t - \phi_1) + M_2 \cos(2\omega t - \phi_2) + \cdots \quad (7.5)$$

with

$$M = \sqrt{A^2 + B^2} \quad \tan \phi = \frac{A}{B} \quad (7.6)$$

The resultant M , the same amplitude as in (7.4), now lags ϕ radians behind the B vector, Fig. 7.1.

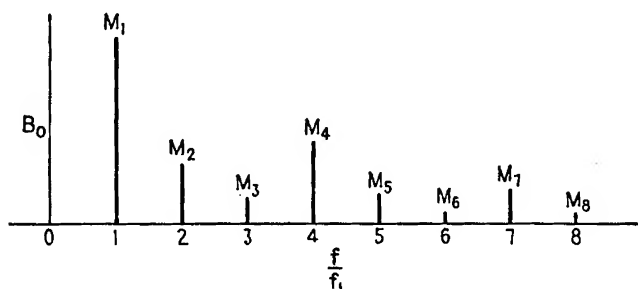


FIG. 7.2.—Example of an amplitude spectrum.

Note that in (7.4) the constants $\psi_1, \psi_2, \psi_3, \dots$ are not the phases of the successive components of different frequencies with respect to one another, since the phase difference between two components of different frequencies or the angle between two vectors rotating at different angular velocities is a function of time. ψ_1 is the phase angle of the fundamental $M_1 \sin(\omega t + \psi_1)$ with respect to $\sin \omega t$, arbitrarily chosen as the standard of reference; ψ_2 is the phase angle of the second harmonic $M_2 \sin(2\omega t + \psi_2)$ with respect to $\sin 2\omega t$ taken as standard for the second harmonic; etc. In (7.5) the cosines are taken as standard. In a graph of voltage plotted against time, each successive component occupies a definite position with respect to the given periodic curve, irrespective of the arbitrarily chosen origin of time.

The values B_0, M_1, M_2, \dots of the amplitude of the successive components constitute the Fourier spectrum of $e(t)$. Certain devices are capable of selecting one frequency at a time and measuring its amplitude. They are called wave analyzers.

A plot of the successive amplitudes as a function of frequency is called the amplitude spectrum of the wave. An example of a

spectrum plotted in terms of f/f_1 is given in Fig. 7.2. Each amplitude M_n is not necessarily smaller than the preceding one (here $M_4 > M_3$, $M_7 > M_6$), but M_n eventually becomes smaller and smaller as n increases.

A knowledge of all the successive amplitudes is necessary but not sufficient to reconstruct the periodic wave $e(t)$ since all the phase

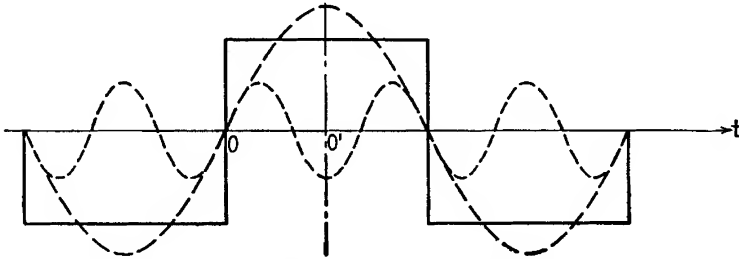


FIG. 7.3.—Square wave, together with its fundamental and third-harmonic components.

angles are required to establish the position of each harmonic component with respect to the resultant periodic wave.

As previously stated, the amplitudes and the positions of the several harmonic components with regard to the resultant periodic wave do not depend upon the arbitrarily chosen origin of time. For example, the square wave, Fig. 7.3, is a square-sine wave if the origin of time is at O and a square-cosine wave if the origin of time is at O' . The Fourier development with the origin at O is (10.1); with the origin at O' is (10.2). These trigonometric expressions are different in form, but each places the various harmonic components in the same position with respect to the original square wave, as shown in Fig. 7.3 for the fundamental and third harmonic.

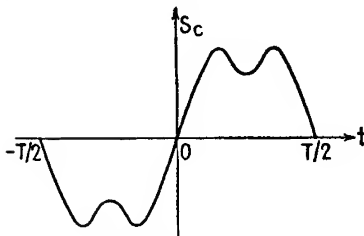


FIG. 8.1.— S function having C symmetry during half periods.

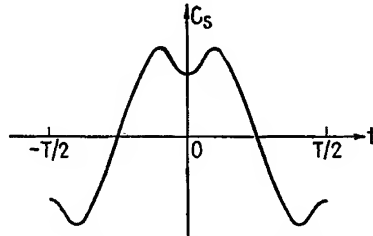


FIG. 8.2.— C function having S symmetry during half periods.

8. Quarter-wave Symmetries.—It often happens that an S function considered over a half period exhibits a C symmetry,

Fig. 8.1, or vice versa, Fig. 8.2. Every other coefficient in the development of the S or C function, (5.1) or (4.1), is then zero:

$$S_C(t) = A_1 \sin \omega t + A_3 \sin 3\omega t + A_5 \sin 5\omega t + \cdots \quad (8.1)$$

$$C_S(t) = B_1 \cos \omega t + B_3 \cos 3\omega t + B_5 \cos 5\omega t + \cdots \quad (8.2)$$

Equations (8.1), (8.2) represent the same wave, the only difference being that the origin of time, $t = 0$, is at a different point of symmetry in (8.1) from that for (8.2).

These quarter-wave symmetrical functions contain therefore no even harmonics. They are particular cases of waves described in Sec. 9.

9. Absence of Even Harmonics.—A periodic wave may present neither symmetry nor antisymmetry and yet have identically shaped positive and negative half periods, Fig. 9.1. Algebraically, $e(t + T/2) = -e(t)$.

The functions $\sin (2k + 1)\omega t$ and $\cos (2k + 1)\omega t$ satisfy this relation, while $\sin 2k\omega t$ and $\cos 2k\omega t$ do not. Hence, a function

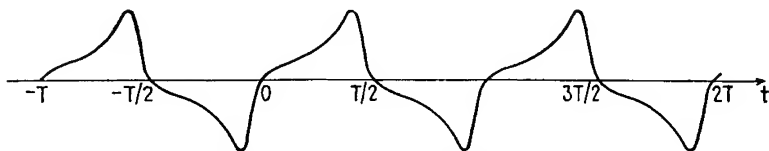


FIG. 9.1.—Periodic function containing no even harmonics.

$e(t)$ having identically shaped positive and negative half periods cannot contain components of the form $\sin 2k\omega t$ or $\cos 2k\omega t$, and therefore can contain no even harmonics. Then

$$e(t) = M_1 \sin (\omega t + \psi_1) + M_3 \sin (3\omega t + \psi_3) + M_5 \sin (5\omega t + \psi_5) + \cdots \quad (9.1)$$

or

$$e(t) = A_1 \sin \omega t + A_3 \sin 3\omega t + A_5 \sin 5\omega t + \cdots + B_1 \cos \omega t + B_3 \cos 3\omega t + B_5 \cos 5\omega t + \cdots \quad (9.2)$$

10. Examples of Fourier Developments.—In the following examples, the indicated peak values of the waveforms have been standardized to the value unity. If, for example, the peak value of an actual voltage is \hat{E} volts, the Fourier development given here should be multiplied by \hat{E} . Each Fourier development and the corresponding figure bear the same number.

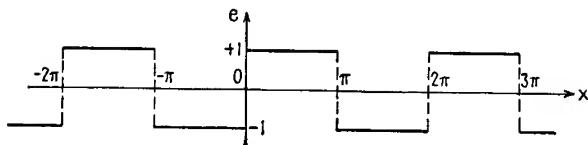


FIG. 10.1.—Square-sine.

$$\begin{aligned} \text{Square-sine } (x) = \frac{4}{\pi} & \left[\sin x + \frac{1}{3} \sin 3x + \frac{1}{5} \sin 5x + \cdots \right. \\ & \left. + \frac{1}{2n+1} \sin (2n+1)x + \cdots \right] \end{aligned} \quad (10.1)$$

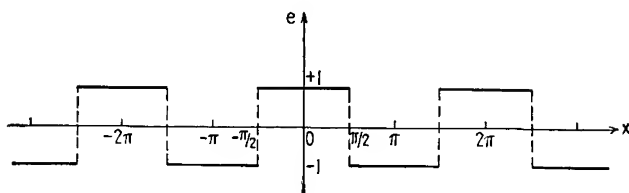


FIG. 10.2.—Square-cosine.

$$\begin{aligned} \text{Square-cosine } (x) = \frac{4}{\pi} & \left[\cos x - \frac{1}{3} \cos 3x + \frac{1}{5} \cos 5x - \frac{1}{7} \cos 7x + \cdots \right. \\ & \left. + \frac{(-1)^n}{2n+1} \cos (2n+1)x + \cdots \right] \end{aligned} \quad (10.2)$$

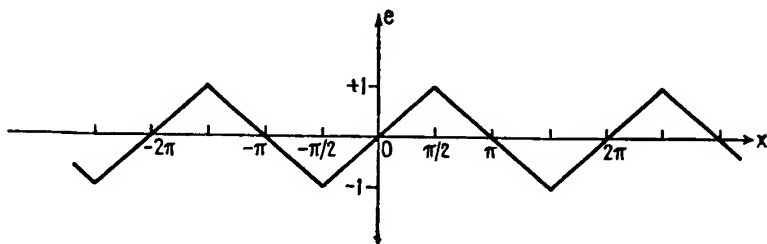


FIG. 10.3.—Antisymmetrical sawtooth.

$$\begin{aligned} e(x) = \frac{8}{\pi^2} & \left[\sin x - \frac{1}{9} \sin 3x + \frac{1}{25} \sin 5x - \frac{1}{49} \sin 7x + \cdots \right. \\ & \left. + \frac{(-1)^n}{(2n+1)^2} \sin (2n+1)x + \cdots \right] \end{aligned} \quad (10.3)$$

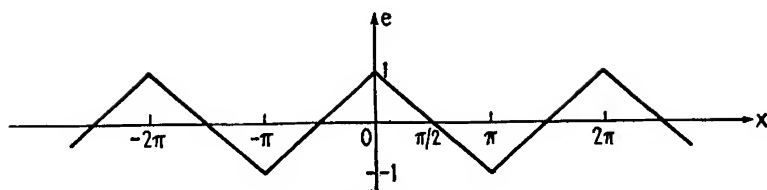


FIG. 10.4.—Symmetrical sawtooth.

$$e(x) = \frac{8}{\pi^2} \left[\cos x + \frac{1}{9} \cos 3x + \frac{1}{25} \cos 5x + \frac{1}{49} \cos 7x + \cdots \right. \\ \left. + \frac{1}{(2n+1)^2} \cos (2n+1)x + \cdots \right] \quad (10.4)$$

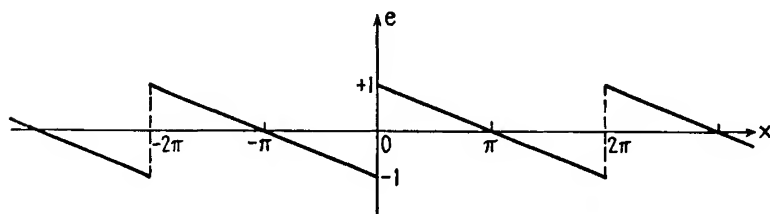


FIG. 10.5.—One example of a scanning voltage.

$$e(x) = \frac{2}{\pi} \left[\sin x + \frac{1}{2} \sin 2x + \frac{1}{3} \sin 3x + \frac{1}{4} \sin 4x + \cdots \right. \\ \left. + \frac{1}{n} \sin nx + \cdots \right] \quad (10.5)$$

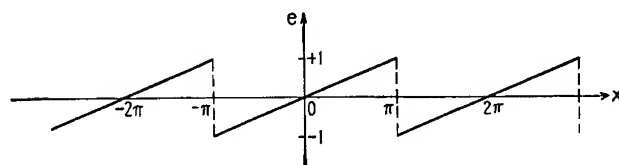


FIG. 10.6.—Another example of a scanning voltage.

$$e(x) = \frac{2}{\pi} \left[\sin x - \frac{1}{2} \sin 2x + \frac{1}{3} \sin 3x - \frac{1}{4} \sin 4x + \cdots \right. \\ \left. + \frac{(-1)^{n+1}}{n} \sin nx + \cdots \right] \quad (10.6)$$

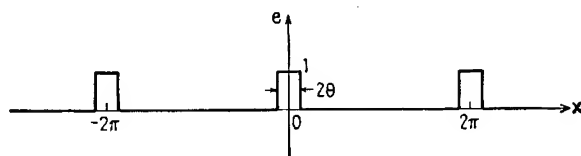


FIG. 10.7.—Periodic rectangular pulse.

$$e(x) = \frac{\theta}{\pi} + \frac{2}{\pi} \left[\sin \theta \cos x + \frac{\sin 2\theta}{2} \cos 2x + \frac{\sin 3\theta}{3} \cos 3x + \cdots + \frac{\sin n\theta}{n} \cos nx + \cdots \right] \quad (10.7)$$

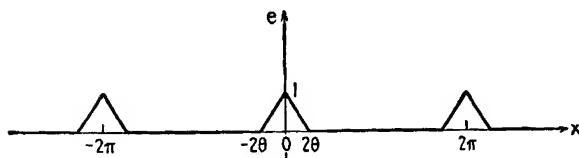


FIG. 10.8.—Periodic triangular pulse

$$e(x) = \frac{\theta}{\pi} + \frac{2}{\pi\theta} \left[\sin^2 \theta \cos x + \frac{\sin^2 2\theta}{4} \cos 2x + \frac{\sin^2 3\theta}{9} \cos 3x + \cdots + \left(\frac{\sin n\theta}{n} \right)^2 \cos nx + \cdots \right] \quad (10.8)$$

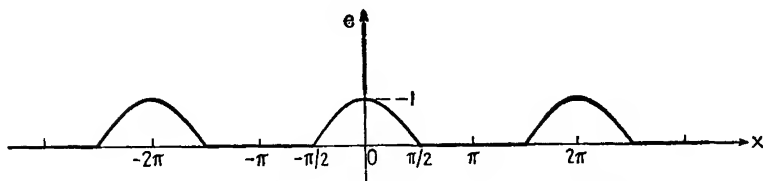


FIG. 10.9.—Output of single-phase half-wave rectifier.

$$e(x) = \frac{1}{\pi} + \frac{1}{2} \cos x + \frac{2}{\pi} \left[\frac{1}{3} \cos 2x - \frac{1}{15} \cos 4x + \frac{1}{35} \cos 6x - \cdots + \frac{(-1)^{n+1}}{4n^2 - 1} \cos 2nx + \cdots \right] \quad (10.9)$$

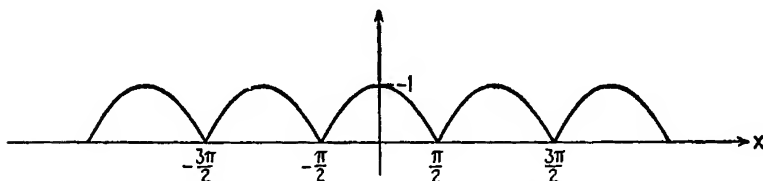


FIG. 10.10.—Output of single-phase full-wave rectifier.

$$e(x) = \frac{2}{\pi} + \frac{4}{\pi} \left[\frac{1}{3} \cos 2x - \frac{1}{15} \cos 4x + \frac{1}{35} \cos 6x - \cdots + \frac{(-1)^{n+1}}{4n^2 - 1} \cos 2nx + \cdots \right] \quad (10.10)$$

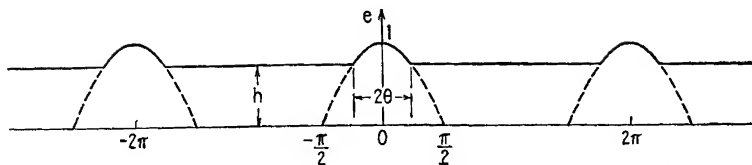


FIG. 10.11.—Portion of a sinusoid of unit amplitude exceeding a height $h = \cos \theta$.

$$e(x) = B_0 + B_1 \cos x + \cdots + E_n \cos nx + \cdots \quad (10.11)$$

$$B_0 = \frac{1}{\pi} [\sin \theta + (\pi - \theta) \cos \theta] \quad (10.12)$$

$$B_n = \frac{1}{\pi n} \left[\frac{\sin (n-1)\theta}{n-1} - \frac{\sin (n+1)\theta}{n+1} \right] \quad (10.13)$$

11. Waves Having Discontinuities.—In Fig. 10.1 the square-sine wave at $x = 0$ jumps from the value -1 to the value $+1$. This function is discontinuous for the value $x = 0$. Similarly, the square-cosine wave is discontinuous for $x = \pi/2$. In such a case the amplitudes of the successive components diminish very slowly, not always exactly as $1/n$, although the decrease is of the order of $1/n$. Other examples are given in Figs. 10.5 to 10.7.

When $e(x)$ jumps discontinuously from a to b for a certain value of x , the sum of all the terms of its Fourier development, for the value of x at which the jump occurs, is $(a+b)/2$. For example, square-sine x is 0 for $x = 0$, since $a = -1$ and $b = +1$.

The sum of a finite number of terms in the Fourier series for a function having a jump is only a fair approximation to the value of the function in the vicinity of the jump.

Other types of waves present no jumps but have sharp corners, such as the waves of Figs. 10.3, 10.4, 10.8, 10.9, 10.10. These waves are continuous, but their first derivatives are discontinuous and are characterized by jumps similar to those of the square-sine and square-cosine functions. Though for such waves the successive amplitudes of the Fourier components decrease slowly, the decrease is more rapid than for the square sine or square cosine, being of the order of $1/n^2$.

The smoother the wave, the higher the order of the derivative at which a jump first occurs. The amplitudes of the successive components diminish more rapidly, of the order of $1/n^{k+1}$, where k is the order of the derivative at which a jump first occurs.

If all the derivatives of a periodic function are continuous, the amplitudes of its Fourier components decrease very rapidly and the expression for M contains as a rule some numerical fraction

raised to the n th power. Vacuum-tube oscillators with a slight positive feedback give voltages of this type. Relaxation oscillators, such as multivibrators or thyatron oscillators, give voltages presenting jumps or kinks. In waves having jumps or kinks the harmonic components of high frequency have appreciable amplitudes.

12. Use of Fourier Series in Transmission Problems.—It is stated in Sec. 2 that the solution of a transmission problem involving periodic nonharmonic voltages comprises three stages: (1) Fourier analysis of input voltage or current; (2) calculations on the network, according to a-c theory, for each component frequency; (3) Fourier synthesis of the output.

Suppose that a square-wave voltage $e_1(t)$ of amplitude E_1 volts and period T is applied to the terminals 1-2 of the network of Fig.

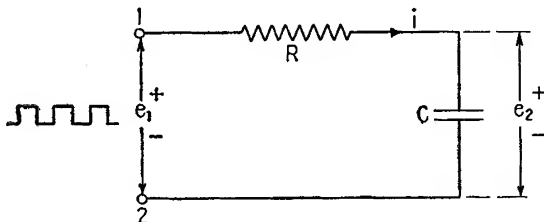


FIG. 12.1.—Example of a transmission problem.

12.1. The problem is to find the final periodic potential difference across the capacitor C after the square wave has been applied for a long time, *i.e.*, after the steady state has been reached.

For harmonic voltages of any frequency ω

$$E_1 = \left(R - \frac{j}{\omega C} \right) I \quad E_2 = -j \frac{1}{\omega C} I \quad (12.1)$$

Therefore,

$$\frac{E_2}{E_1} = A = \frac{1}{1 + jR\omega C} \quad (12.2)$$

This complex *voltage ratio* can be written in the equivalent forms

$$\frac{E_2}{E_1} = A = |A| \angle -\beta = e^{-\alpha - j\beta} \quad (12.3)$$

The numerics α and β are called the *attenuation* and the *phase lag* of the network. α is measured in *nepers*, β in *radians*. If α and β are both *positive*, say $\alpha = 1$ neper, $\beta = \pi/6$, Fig. 12.2,

the vector E_2 will be *smaller* than the vector E_1 in the ratio $1/2.7$ and will *lag* the vector E_1 by $\pi/6$, or 30° .

The numerics A , α , β will vary in general with frequency. In the present example,

$$|A| = \epsilon^{-\alpha} = \frac{1}{\sqrt{1 + R^2 \omega^2 C^2}} \quad (12.4)$$

so that $|A|$ decreases from 1 to zero. The attenuation α will increase from zero to infinity. From (12.2), the vector E_2 is lagging E_1 . The phase lag β is *positive*, increasing from zero to $\pi/2$, Fig. 12.3. The locus of $A = \epsilon^{-\alpha - j\beta}$ is a half circle, Fig. 12.4.

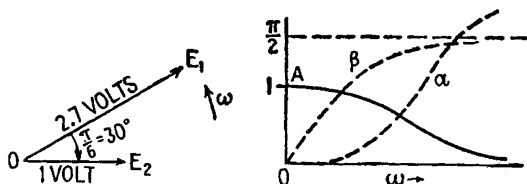


FIG. 12.2.—Vector diagram illustrating an attenuation of one neper and a phase lag of $\pi/6$ radians.

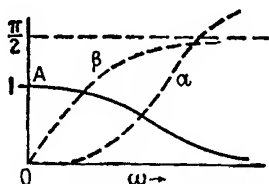


FIG. 12.3.—Frequency characteristics of the network of Fig. 12.1.

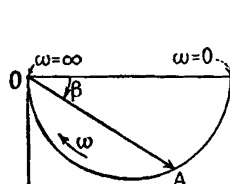


FIG. 12.4.—Locus of the complex voltage ratio $A = \epsilon^{-\alpha - j\beta}$.

The quantity

$$\omega_0 = \frac{1}{CR} \quad (12.5)$$

will be taken as a reference angular frequency; A and $\tan \beta$ can then be written as

$$A = \frac{1}{\sqrt{1 + \frac{\omega^2}{\omega_0^2}}} \quad (12.6)$$

$$\tan \beta = \frac{\omega}{\omega_0} \quad (12.7)$$

The given input voltage (square wave) has the Fourier development, from (10.1),

$$e_1(t) = \frac{4\hat{E}_1}{\pi} (\sin \omega t + \frac{1}{3} \sin 3\omega t + \frac{1}{5} \sin 5\omega t + \dots) \quad (12.8)$$

The fundamental component will come out of the network with its amplitude multiplied by A_1 and its phase angle diminished by β_1 , A_1 and β_1 being obtained from (12.6) and (12.7). The third harmonic will come out multiplied by A_3 and the phase diminished by

β_3 , A_3 and β_3 being obtained by replacing in (12.6) and (12.7) ω by 3ω , etc. The sum of all these harmonic components will be the output voltage $e_2(t)$.

In general, the summation of these components (Fourier synthesis) is difficult to work out. Two particular cases will be considered.

13. Example of an Integrator.—Assume the angular frequency ω of the applied square wave to be large with respect to the reference angular frequency $\omega_0 = 1/CR$,

$$\omega \gg \omega_0 \quad \text{or} \quad T \ll CR \quad (13.1)$$

Then, approximately,

$$A_1 \doteq \frac{\omega_0}{\omega} \quad (13.2)$$

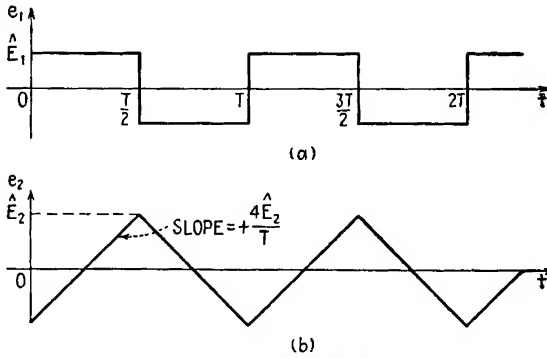


FIG. 13.1.—Input and output voltages of an integrator, showing phase relationship.

and, to a better and better approximation,

$$A_3 \doteq \frac{\omega_0}{3\omega} \quad A_5 \doteq \frac{\omega_0}{5\omega} \quad \dots \quad (13.3)$$

The phase lags β_1 , β_3 , β_5 , . . . will be approximately $\pi/2$. Each sine wave in (12.8) will have its phase angle *decreased* by $\pi/2$, which changes it to $(-\cos)$, and the successive output components will be

$$-\frac{4\hat{E}_1}{\pi} \frac{\omega_0}{\omega} \cos \omega t - \frac{4\hat{E}_1}{3\pi} \frac{\omega_0}{3\omega} \cos 3\omega t - \frac{4\hat{E}_1}{5\pi} \frac{\omega_0}{5\omega} \cos 5\omega t \dots$$

and

$$e_2(t) = -\frac{4\hat{E}_1}{\pi} \frac{\omega_0}{\omega} (\cos \omega t + \frac{1}{9} \cos 3\omega t + \frac{1}{25} \cos 5\omega t + \dots) \quad (13.4)$$

From (10.4) this is the development of a symmetrical saw tooth, Fig. 13.1b. The common factor in (13.4) can be written as

$$\frac{4\hat{E}_1}{\pi} \frac{1}{CR} \frac{T}{2\pi} = \frac{2}{\pi^2} \frac{T}{CR} \hat{E}_1 \quad (13.5)$$

The common factor in (10.4) is $8/\pi^2$ for a peak amplitude 1 and $(8/\pi^2)\hat{E}_2$ for the peak amplitude \hat{E}_2 of the output voltage. Equating this common factor to the common factor of (13.5),

$$\frac{8}{\pi^2} \hat{E}_2 = \frac{2}{\pi^2} \frac{T}{CR} \hat{E}_1 \quad \text{or} \quad \hat{E}_2 = \frac{T}{4CR} \hat{E}_1 \quad (13.6)$$

and, from (13.1), \hat{E}_2 is very much less than \hat{E}_1 .

Note that $e_1(t)$ is the derivative of $e_2(t)$ times a constant CR . Therefore $e_2(t)$ is an *integral* of $e_1(t)$, and when $CR \gg T$ the network acts approximately as an *integrator*.

14. Example of Negligible Distortion.—Assume that T is large with respect to CR , or $\omega \ll \omega_0$. From (12.6),

$$A_1 \doteq A_3 \doteq A_5 \cdots \doteq 1,$$

to the point where

$$n^2\omega^2 < \omega_0^2 \quad (14.1)$$

no longer holds. Call m the order of the harmonic beyond which (14.1) ceases to hold. From (12.7),

$$\beta_1 \doteq \frac{\omega}{\omega_0} \quad \beta_3 = \frac{3\omega}{\omega_0} \quad \beta_5 = \frac{5\omega}{\omega_0} \quad \cdots \quad (14.2)$$

for the first m harmonics.

The idea of time lag, or *time delay*, must now be introduced. Consider two harmonic voltages, or waves, e_1 and e_2 of the same angular frequency ω and period T , Fig. 14.1. Their amplitudes do not matter. In the figure the voltage wave e_2 lags e_1 by one-quarter period, or $T/4$ sec; the rotating vector representing E_2 lags the vector E_1 by 90° , or $\pi/2$ radians. In general, if the vector E_2 lags E_1 by β radians, the wave e_2 will lag e_1 by t_d sec, and the relation between time delay t_d and phase lag β will be

$$t_d = \frac{T}{2\pi} \beta = \frac{\beta}{\omega} \quad (14.3)$$

The time delay of the successive components of the output voltage with respect to the input component voltage of the same

frequency is, from (14.2),

$$t_{d1} = \frac{\beta_1}{\omega} \doteq \frac{\omega}{\omega_0 \omega} \doteq \frac{1}{\omega_0} \quad t_{d3} = \frac{\beta_3}{3\omega} \doteq \frac{1}{\omega_0} \quad \dots \quad (14.4)$$

or the time delay for the first m harmonics is approximately the same, $t_d \doteq 1/\omega_0 = CR$ sec. Thus the first m harmonics of the input voltage come out of the network with their amplitudes unchanged ($A \doteq 1$) and are delayed by the same amount, CR sec. Their sum is a periodic nonharmonic output voltage of exactly the same shape as the sum of the first m harmonics of the input voltage. If the harmonics of order higher than m can be neglected, the network, when $CR \gg T$, transmits the input voltage with negligible distortion but with time delay.

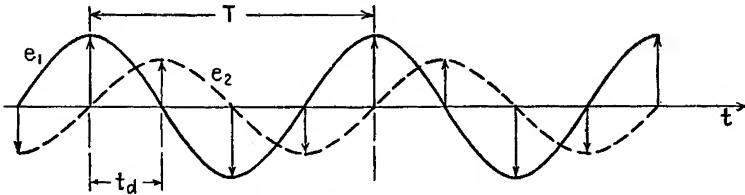


FIG. 14.1.—Illustrating time delay between two waves.

15. Conditions of Nondistortion.—This last result can be generalized. If for a given network the magnitude of the voltage ratio, $|A| = \epsilon^{-\alpha}$, is independent of frequency, and if the time delay t_d is also independent of frequency, every harmonic component of a periodic input voltage will be magnified $|A|$ times and delayed t_d sec so that the periodic output voltage will be obtained by magnifying the input voltage $|A|$ times and delaying it t_d sec. The following *two conditions* are therefore *sufficient* to ensure nondistortion:

$$\left. \begin{array}{l} (1) |A| \text{ constant with regard to frequency} \\ (2) t_d \text{ constant with regard to frequency} \end{array} \right] \quad (15.1)$$

These conditions can be expressed also in terms of the attenuation α and phase lag β , as follows:

$$\left. \begin{array}{l} (1) \alpha \text{ constant with regard to frequency} \\ (2) \beta \text{ proportional to frequency } (\beta = t_d \omega) \end{array} \right] \quad (15.2)$$

$|A|$ or α equal to a constant is in any case a *necessary* condition for nondistortion. As regards β there are, however, other possibilities than that expressed by $\beta = t_d \omega$. First, if $\beta = \pi$ radians for

all frequencies, every output harmonic will be in phase opposition to the corresponding input harmonic, so that the output periodic wave will have the same shape as the input periodic wave, only inverted with respect to the time axis. Second, the condition for this inverted wave to be delayed by t_d sec is

$$\beta = \pi + t_d\omega \quad (15.3)$$

The phase-lag condition of (15.2) is represented by the straight line 1, the condition of (15.3) by line 2, Fig. 15.1. It is clear that $\beta = 2\pi$ or $\beta = 2\pi + t_d\omega$ (line 3, Fig. 15.1) would also be a suitable condition. Line 2, Fig. 15.1, is important as representing the conditions approximated in a one-stage amplifier (Chap. XIII).

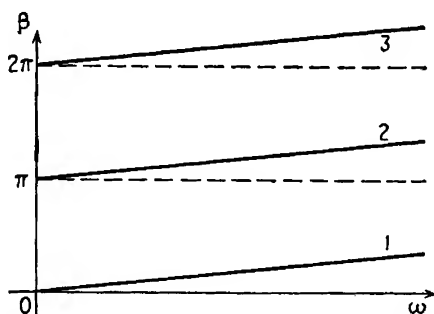


FIG. 15.1.—Different types of variation of phase lag β , none of which introduces phase distortion.

Often the conditions (15.1) or (15.2) for $|A|$ and β will not hold rigorously for all the multiples of the fundamental frequency, but they may hold approximately for the first m harmonics, Sec. 14, in which case the distortion of the input voltage wave will usually be negligible.

In general, if the attenuation α is constant with regard to frequency and if the phase lag of the periodic output voltage $e_2(t)$ with regard to the periodic input voltage $e_1(t)$ is given by

$$\beta = n\pi + t_d\omega \quad (n = 0, \pm 1, \pm 2, \dots) \quad (15.4)$$

the voltage $e_2(t)$ will be $\epsilon^{-\alpha}$ as large as the input voltage and delayed by t_d sec when $n = 0, \pm 2, \dots$. When $n = \pm 1, \pm 3, \dots$, the output voltage will be $\epsilon^{-\alpha}$ as large as the input voltage, delayed by t_d sec and inverted with respect to the time axis. A network producing a time delay is called a *delay network*.

Mathematically, the above would be equally true if t_d were

negative, but physically it is impossible to construct a network that will advance a signal in time instead of delaying it.

Distortion that occurs when $|A|$ or α is not constant with respect to frequency is called *frequency distortion*. Distortion that occurs when β is not proportional to frequency (or when t_d is not a constant with respect to frequency) is called *phase distortion* (or *delay distortion*).

16. A Numerical Method of Fourier Analysis.—There are several numerical processes (called schedules) for calculating the harmonic components of a complex waveform. The particular analysis to be described was developed¹ to apply especially to vacuum-tube problems though the analysis may be found useful elsewhere.

The analysis applies only when certain conditions hold. Suppose

$$y = f(x) \quad (16.1)$$

represents a nonlinear relation between the dependent variable y and the independent variable x . For example, y may be the plate current of a vacuum tube as determined by the grid voltage e_g or the plate voltage e_p , either of which may be represented by x . Or y may be the power determined by some variable x such as voltage or current. Generally, (16.1) is given in graphical form, for example, as the path of operation on a characteristic chart of a vacuum tube. The conditions under which the analysis can be applied are

1. The variable y must be a single-valued function of x .
2. The variable x must vary sinusoidally with time.

These conditions hold in the example shown in Fig. 16.1, where $i = f(e)$ represents an e - i characteristic of some device. Condition 1 is met since the closed path of the representative point (e, i) is the double line $A'QA''QA'$, or the current i has the same value for a given e whether the voltage e is increasing or decreasing. Consequently, the waveform of y as a function of time will have half-wave symmetry. Condition 2 is met since the impressed voltage e is equal to $\bar{E} + \hat{E} \cos \omega t$.

In what follows, e and i will be measured from the point Q taken as a new origin. The problem is to find the average current and the harmonic components of the current that flows when the

¹ E. L. CHAFFEE, "A Simplified Harmonic Analysis," *Rev. Sci. Inst.*, **7**, 384, October, 1936.

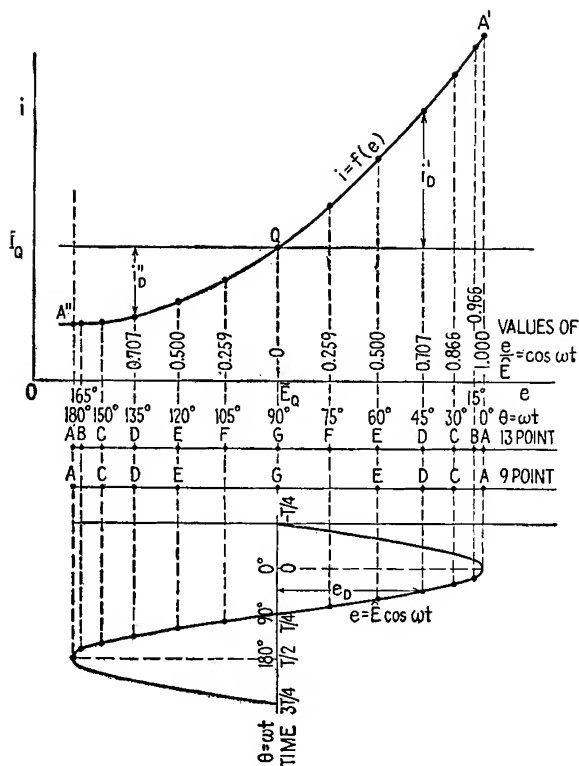


FIG. 16.1.—Current-voltage relationship and points used in harmonic analysis.

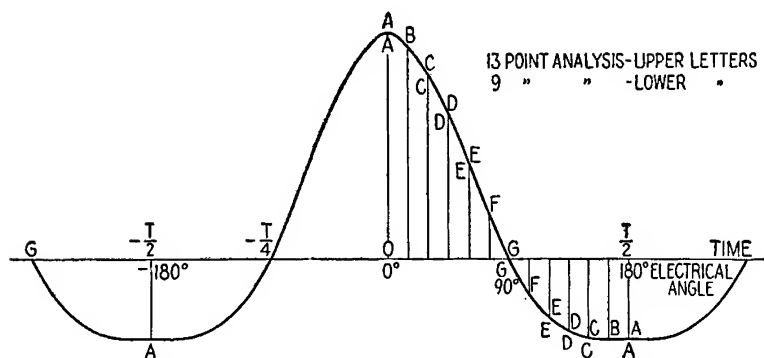


FIG. 16.2.—Waveform of current under the conditions of Fig. 16.1.

impressed voltage is $\cos \hat{E} \omega t$, represented by the sinusoidal waveform in the lower part of Fig. 16.1.

The current waveform is the distorted sinusoid in Fig. 16.2, of which $1\frac{1}{2}$ periods are shown. The Fourier analysis of this waveform is based on the values of i for certain specified points on the curve. Each point corresponds to a specified value of the electrical angle $\theta = \omega t$ and therefore to a specified value of $e/\hat{E} = \cos \theta$. It is unnecessary, however, to draw the actual waveform of Fig. 16.2, since the desired values of i can be read off directly from the e - i characteristic of Fig. 16.1.

The number of specified values of θ or of e/\hat{E} depends upon the irregularity of the characteristic and upon the accuracy of analysis desired. Only two analyses will be described, one making use of 9 points and the other making use of 13 points of one half cycle.

The 13 values of θ and e/\hat{E} for the 13-point analysis and the 9 values for the 9-point analysis are indicated in Fig. 16.1 and are given in Table 16.1.

TABLE 16.1

Electrical angle $\theta = \omega t$	Values of $e/\hat{E} = \cos \theta$	Points for 13-point analysis	Points for 9-point analysis
0°, 180°	± 1	<i>A</i>	<i>A</i>
15°, 165°	± 0.966	<i>B</i>	Not used
30°, 150°	± 0.866	<i>C</i>	<i>C</i>
45°, 135°	± 0.707	<i>D</i>	<i>D</i>
60°, 120°	± 0.500	<i>E</i>	<i>E</i>
75°, 105°	± 0.259	<i>F</i>	Not used
90°	0	<i>G</i>	<i>G</i>

The first step is to read from the curve the two values of i for each pair of points identified by each capital letter, Fig. 16.1. For example, determine i_D' and i_D'' , corresponding to $e/\hat{E} = \pm 0.707$. Then find the algebraic sum and difference of i_D' and i_D'' ,

$$\begin{aligned} P_D &= i_D' + i_D'' \\ M_D &= i_D' - i_D'' \end{aligned} \quad (16.2)$$

(The letters P and M are the initial letters of plus and minus.)

The next step is to fill in the various values of P and M corresponding to the appropriate letters in Table 16.2 for the 9-point analysis or in Table 16.3 for the more accurate 13-point analysis.

TABLE 16.2.—9-POINT ANALYSIS

$$\begin{aligned}\bar{I} - \bar{I}_Q &= \frac{1}{6} \left[\frac{P_A}{2} + P_C + P_E \right] \\ (\hat{I})_1 &= \frac{1}{6} [M_A + \sqrt{3} M_C + M_E] \\ (\hat{I})_2 &= \frac{1}{6} [P_A + P_C - P_E] \\ (\hat{I})_3 &= \frac{1}{6} [M_A - 2M_E] \\ (\hat{I})_4 &= \frac{1}{6} [P_A - 2P_D] \\ (\hat{I})_5 &= \frac{1}{12} [M_A - 3\sqrt{2} M_D + 4M_E] \\ (\hat{I})_6 &= \frac{1}{12} [P_A - 2P_C + 2P_E] \\ (\hat{I})_7 &= \frac{1}{12} [M_A - 2\sqrt{3} M_C + 3\sqrt{2} M_D - 2M_E] \\ (\hat{I})_8 &= \frac{1}{12} \left[\frac{P_A}{2} - 2P_C + 3P_D - 2P_E \right]\end{aligned}$$

TABLE 16.3.—13-POINT ANALYSIS

$$\begin{aligned}\bar{I} - \bar{I}_Q &= \frac{1}{12} \left[\frac{P_A}{2} + P_B + P_C + P_D + P_E + P_F \right] \\ (\hat{I})_1 &= \frac{1}{12} [M_A + 1.932M_B + 1.732M_C + 1.414M_D + M_E + 0.518M_F] \\ (\hat{I})_2 &= \frac{1}{12} [P_A + 1.732P_B + P_C - P_E - 1.732P_F] \\ (\hat{I})_3 &= \frac{1}{12} [M_A - 1.414M_B - 1.414M_D - 2M_E - 1.414M_F] \\ (\hat{I})_4 &= \frac{1}{12} [P_A + P_B - P_C - 2P_D - P_E + P_F] \\ (\hat{I})_5 &= \frac{1}{12} [M_A + 0.518M_B - 1.732M_C - 1.414M_D + M_E + 1.932M_F] \\ (\hat{I})_6 &= \frac{1}{12} [P_A - 2P_C + 2P_E] \\ (\hat{I})_7 &= \frac{1}{12} [M_A - 0.518M_B - 1.732M_C + 1.414M_D + M_E - 1.932M_F] \\ (\hat{I})_8 &= \frac{1}{12} [P_A - P_B - P_C + 2P_D - P_E - P_F] \\ (\hat{I})_9 &= \frac{1}{12} [M_A - 1.414M_B + 1.414M_D - 2M_E + 1.414M_F] \\ (\hat{I})_{10} &= \frac{1}{12} [P_A - 1.732P_B + P_C - P_E + 1.732P_F] \\ (\hat{I})_{11} &= \frac{1}{12} [M_A - 1.932M_B + 1.732M_C - 1.414M_D + M_E - 0.518M_F] \\ (\hat{I})_{12} &= \frac{1}{12} \left[\frac{P_A}{2} - P_B + P_C - P_D + P_E - P_F \right]\end{aligned}$$

17. Analysis of a Simple Specific Case.—It may happen in practice, for example in Class B and Class C amplifiers, Chaps.

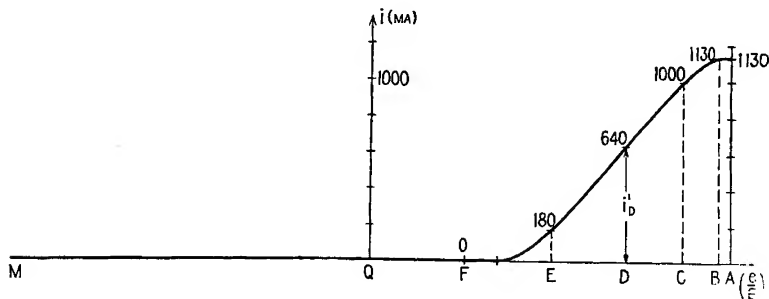


FIG. 17.1.—Current-voltage relationship when there is no current for negative values of e ; all currents $i'' = 0$.

XIII and XIV, that the variable y of (16.1) is zero for all negative values of x at least. Figure 17.1 illustrates the e - i characteristic

in an actual case, where the current begins to rise at $e/\hat{E} \doteq 0.35$. The quiescent current is zero, and all the currents i'' of Figs. 16.1 and 16.2 are zero. Then $P = M = i''$ at all points. The formulas for the average and for the fundamental components of Tables 16.2 and 16.3 reduce to the following:

TABLE 17.1.—9-POINT ANALYSIS
(All currents $i'' = 0$)

$$\bar{I} = \frac{1}{6} \left[\frac{i_A}{2} + i_C + i_E \right]$$

$$(\hat{I})_1 = \frac{1}{6} [i_A + \sqrt{3} i_C + i_E]$$

Note that, if only the d-c and fundamental components are desired, only six quantities need be measured in general for the 9-point analysis, Table 16.2; and if all currents i'' are zero, only three quantities are needed.

TABLE 17.2.—13-POINT ANALYSIS
(All currents $i'' = 0$)

$$\bar{I} = \frac{1}{12} \left[\frac{i_A}{2} + i_B + i_C + i_D + i_E + i_F \right]$$

$$(\hat{I})_1 = \frac{1}{12} [i_A + 1.932i_B + 1.732i_C + 1.414i_D + i_E + 0.518i_F]$$

In the numerical example of Fig. 17.1, the results are

9-point analysis: $\bar{I} = 291$ ma, $\hat{I}_1 = 507$ ma

13-point analysis: $\bar{I} = 293$ ma, $\hat{I}_1 = 511$ ma

The 13-point analysis is in general more accurate than the 9-point analysis. The use of \bar{I} and \hat{I}_1 in calculating the power relationships in vacuum tubes is explained in Chap. XIV.

Supplementary Reading

- R. M. KERCHNER and G. F. CORCORAN: "Alternating Current Circuits," Chap. VI, John Wiley & Sons, Inc., New York, 1st ed., 1938, 2d ed., 1943.
- I. S. and E. S. SOKOLNIKOFF: "Higher Mathematics for Engineers and Physicists," 2d ed., Chap. II, McGraw-Hill Book Company, Inc., New York, 1941.
- T. E. SHEA: "Transmission Networks and Wave Filters," Chap. X, D. Van Nostrand Company, Inc., New York, 1929.
- E. A. GUILLEMIN: "Communication Networks," Vol. I, Chap. X, John Wiley & Sons, Inc., New York, 1931.
- A. R. KNIGHT and G. H. FETT: "Introduction to Circuit Analysis," pp. 404-437, Harper & Brothers, New York, 1943.
- R. R. LAWRENCE: "Principles of Alternating Currents," 2d ed., pp. 100-123, McGraw-Hill Book Company, Inc., New York, 1935.
- I. A. PIPES: "Applied Mathematics for Engineers and Physicists, Chap. III, McGraw-Hill Book Company, Inc., New York, 1946.

CHAPTER X

ELECTRON EMISSION AND THE DIODE

The operation of vacuum tubes depends primarily upon the movement of electricity in the form of electrons (negatively charged particles) that move through the spaces between the electrodes. These electrons are released into the space within the tube by one or more of the following processes:

1. Thermionic emission
2. Secondary emission
3. Photoelectric emission
4. High-field emission
5. Ionization

These will be studied in turn.

1. Thermionic Emission.¹—Some of the outer electrons associated with the atoms of a metal are so loosely bound that they can circulate freely among the atoms. These are called *free electrons*. A current of electricity in the metal is a drifting of free electrons.

Associated with the atoms (or molecules) of a substance is an energy of vibration called *heat*. The random thermal vibration of the atoms increases with increase in temperature, as does the thermal agitation communicated by the atoms to the free electrons. As a result of thermal agitation, all the free electrons would spill out of the surface of a conductor if it were not for a surface effect that opposes their escape. This surface restraint, or “barrier,” is quantitatively expressed as the increase of the potential energy (loss of kinetic energy) of an electron as it leaves the metal. It is called the *work function*, its value being characteristic of the particular metal.

The work function can be expressed in terms of the electron affinity, which is defined by the relation $\phi = w/e$, where w is the work function, ϕ is the electron affinity (commonly expressed in

¹ E. L. CHAFFEE, “Theory of Thermionic Vacuum Tubes,” pp. 55–56, 95–97, McGraw-Hill Book Company, Inc., 1933; A. L. ALBERT, “Fundamental Electronics and Vacuum Tubes,” pp. 33–47, 53–72, The Macmillan Company, 1938.

volts), and e is the electronic charge. The value of ϕ for a temperature of 0°K is denoted by ϕ_0 .

Among the characteristics that make a metal desirable as a thermionic emitter are low electron affinity and high melting temperature. These characteristics for certain metallic substances are given in Table 1.1.

TABLE 1.1.—VALUES OF ELECTRON AFFINITY AND MELTING TEMPERATURE

Metal	ϕ_0 , volts	Melting temp., $^\circ\text{K}$
Calcium.....	2.24	1083
Molybdenum.....	4.42	2893
Tantalum.....	4.06	3123
Thorium.....	3.35	2118
Tungsten.....	4.52	3643
Thorium on tungsten.....	2.63	
Oxide-coated nickel.....	0.5–1.5	

The thermionic emission of electrons from a metallic surface depends upon the absolute temperature T and the value of ϕ_0 according to

$$i = AT^2\epsilon^{\frac{-11,600\phi_0}{T}} \text{ amperes/cm}^2 \quad (1.1)$$

in which the constant A for some metals, notably tungsten, tantalum, molybdenum, has the numerical value 60.2. The exponent of ϵ in (1.1) is equal to $-w_0/kT$, where w_0 is the work function at absolute zero and k is Boltzmann's constant.

In Fig. 1.1 the thermionic emission per square centimeter for the metal tungsten is plotted against the absolute temperature T .

When thermionic emission occurs at a surface of an isolated electrode, such as one in an evacuated glass envelope, it becomes surrounded by a cloud of electrons that exerts a force of repulsion sufficient to cause those near the surface to reenter the electrode. A condition of equilibrium is established in which electrons are reentering the surface at the same rate at which they are being emitted.

To verify (1.1) experimentally it is necessary to provide an electrode, positive with respect to the emitting surface, that will collect *all* the electrons that are emitted. The potential of this positive electrode must be so high that practically no electrons, once emitted, can reenter the emitting surface.

The three important types of thermionic emitters are pure metals, thoriated-tungsten emitters, oxide-coated emitters.

Pure-metal Emitters.—Tungsten and tantalum, having high melting points, Table 1.1, can be operated at sufficiently high temperatures to give copious emission without excessive evaporation of the metal.

Tungsten is the most commonly used of the pure metallic emitters.

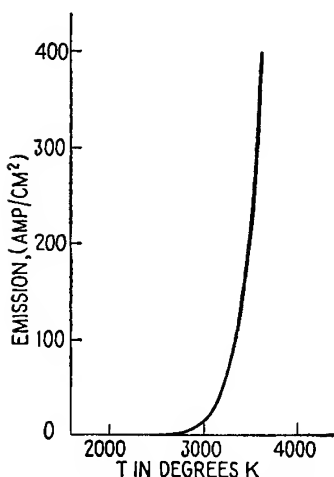


FIG. 1.1.—Thermionic emission from tungsten.

The useful life of a vacuum-tube filament of this type is arbitrarily defined as the operating time (in hours) for which the cross-sectional area of the filament is reduced by 10 per cent. The operating temperature is chosen usually so that the useful life is about 2,000 hr for small tubes, but longer for large power tubes. Representative operating temperatures for tungsten are 2300 to 2700°K. In most cases the voltage applied to the filament, rather than the current, is maintained constant throughout the life of the fila-

ment, since the variation in cross section causes a variation in current (at constant voltage) that maintains a nearly constant operating temperature.

Thoriated-tungsten Emitters.—The presence of a monatomic layer of pure thorium on the surface of a tungsten filament greatly increases its electron emission. If 2 or 3 per cent of thorium oxide is added to a tungsten filament during its manufacture, such a monatomic layer can be secured by any one of several activation processes.

If a thoriated-tungsten filament is used until its emission is greatly reduced, it may be "reactivated" by one of several methods. A typical procedure is to apply 3 to 3½ times the normal filament voltage for 10 to 20 sec (without plate voltage) and, without allowing the filament to cool, to reduce the voltage to 1½ times normal for 1 to 2 hr. As this process is similar to the initial activation process, it serves to renew the monatomic layer of thorium.

Bombardment by gas ions of more than about 22 volts energy rapidly disintegrates the delicate monatomic layer and greatly

decreases emission. If a tungsten filament is heated in hydrocarbon vapor, the outer layer is transformed into tungsten carbide. A monatomic layer of thorium deposited on tungsten carbide by the activation process is less affected by gas-ion bombardment than a layer deposited on pure tungsten. Since a thoriated cathode is very sensitive to residual gas, a "getter" is always employed to maintain a good vacuum in tubes with this type of cathode. The getter is a small amount of chemically active material (magnesium, for example) that is placed in the tube during manufacture. After the tube is evacuated, the getter is fired by the application of heat, which causes it to absorb or combine with practically all the gas left in the tube after the pumping operations. The getter maintains the vacuum after the tube is sealed off by "cleaning up" any gas emitted from the metallic parts of the tube. A getter is also used in tubes having oxide-coated emitters.

Oxide-coated Emitters.—An emitter of the oxide-coated type consists of a metal core coated with an emitting material such as the oxides of the alkaline-earth elements, barium and strontium. Operating temperatures, about 1000°K, are much lower than for pure metals. Oxide-coated emitters must be activated to establish emission.

It is believed that the emission of an oxide-coated emitter is pure thermionic emission from particles of the metal reduced from the oxide, the oxide serving only as a reservoir for the supply of metal. The high emission at low temperatures results from the low work functions of the metals strontium and barium and from the effects of monatomic layers which aid the escape of electrons.

Emission efficiency is defined as milliamperes of emission current per watt of heating power.

TABLE 1.2.—COMPARISON OF THREE TYPES OF EMITTERS

Type of emitter	Operating temp., °K	Emission, ma/watt	Emission, ma/cm ²	Max. plate volts (d-c)
Tungsten.....	2400–2600	2–8	100–1,000	
Oxide-coated.....	1000 approx.	100–300	1,000–3,000	1,000
Thoriated-tungsten.....	1900 approx.	90–100	2,000–3,000	5,000

The oxide-coated type of emitter disintegrates if used continuously at full emission. It is employed in low-power tubes whose plate voltages are less than about 1,000 volts. It is used almost universally in receiving tubes because of its high emission efficiency.

Thoriated-tungsten filaments are used for small power tubes whose normal plate voltages are below about 5,000 volts, while pure tungsten filaments are used only in high-power tubes with plate voltages in excess of about 5,000 volts.

Indirectly Heated Cathodes.—The thermionic emitter, or cathode, of a receiving tube may be a directly heated emitting filament or a coated metal cylinder inside which is an insulated filament that serves only as a heater. The cylinder is thus *equipotential*, so that alternating current can be used in the heater wire without introducing into the circuit an objectionable amount of hum-producing current. The thermal capacity of the cathode prevents appreciable changes of temperature during a cycle of alternating current.

2. Other Types of Emission.—Electron emission can be obtained by other methods. The following types of emission are important in the applications of electronic devices:

Secondary Emission.—An electron moving with the speed acquired by falling through a potential difference greater than 20 or 30 volts can cause the emission of 1 to 10 low-velocity electrons from a metallic surface that it strikes. This “knocking out” of electrons from the surface is called secondary emission, a phenomenon that greatly affects the characteristics of vacuum tubes. Secondary emission is increased markedly by the presence of thin films of metals (barium, thorium, etc.), which are formed on the electrodes of tubes in which these substances are used.

Photoelectric Emission.—When light of proper character falls upon a metallic surface, electrons are emitted from the surface, the energy necessary to remove them being obtained through the absorption of luminous energy. For electrons to be emitted, the frequency of the light must exceed a certain critical value, which is proportional to the work function of the metal. When emission does occur, it is proportional to the intensity of illumination.

High-field Emission.—When the electric field at the surface of a metal is very high, of the order of 1,000,000 volts/cm, electrons may be pulled out through the surface. Undesired emission from the grid of a high-power tube is sometimes due to this cause.

Ionization.—One or more of the loosely bound electrons in an atom of residual gas may be knocked out of the atom by collision with a rapidly moving electron. The products of the collision include one or more free electrons (negative) and a positive ion that is the remaining portion of the atom. As a result of this

ionization, large currents can be produced in tubes that contain considerable amounts of residual gas.

3. Current-Voltage Characteristics of Diodes (Static Characteristics).¹—The current through a diode, called the plate current, is a function of the plate voltage and the temperature of the cathode, or

$$i_b = f(e_b, T) \quad (3.1)$$

where i_b is the current, e_b is the potential of the plate with respect to the cathode, and T is the cathode temperature.

If the plate potential is maintained constant and the current is measured for various values of cathode temperature, the results for a pure-metal cathode are as indicated by one of the family of curves, Fig. 3.1. If the plate potential is increased from E_1 to E_2 or E_3 , the variation of plate current with cathode temperature is shown by the curves so designated. The dotted curve, Fig. 3.1, represents the total thermionic emission. The difference between the emission current and the plate current represents the rate at which electrons return to the cathode after being emitted. At a given value of plate voltage, the plate current builds up as the cathode temperature is increased until the plate current is limited by the space-charge effect, which is due to an accumulation of electrons in the space between the cathode and plate. These electrons exert an electric force that, in the vicinity of the cathode, is opposite to the electric force due to the plate. This action will be described more fully in Sec. 4.

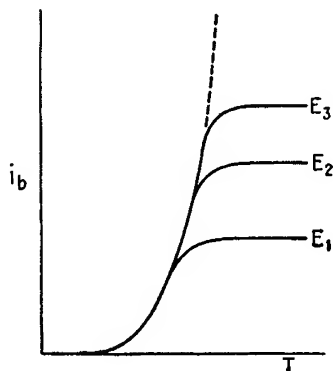


FIG. 3.1.—Plate current versus cathode temperature for a diode.

If the cathode temperature is maintained constant, the variation of current (for a pure-metal cathode) as the plate potential is changed is shown (curve T_1) in Fig. 3.2. With higher temperatures curves T_2 and T_3 result. For a given value of T the limiting value of the plate current is determined by the total thermionic emission

¹ CHAFFEE, *op. cit.*, pp. 66–86; A. V. EASTMAN, “Fundamentals of Vacuum Tubes,” 2d ed., pp. 29–48, McGraw-Hill Book Company, Inc., 1941; H. J. REICH, “Theory and Application of Electron Tubes,” 1st ed., pp. 13–33, 2nd ed., pp. 19–37, McGraw-Hill Book Company, Inc., 1939, 1944.

for that temperature. This limiting current is reached when all the electrons leaving the cathode are drawn to the plate, so that any further increase in plate potential cannot increase the current.

The horizontal portions of the curves of Fig. 3.2 are not representative of the results obtained (in the steady state) for thoriated or oxide-coated cathodes. The surfaces of these emitters are changed when high plate currents are steadily maintained, and their emission characteristics are altered. If the plate voltage

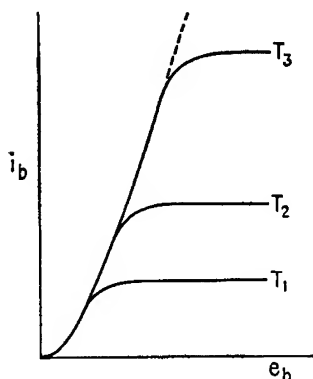


FIG. 3.2.—Plate current versus plate potential.

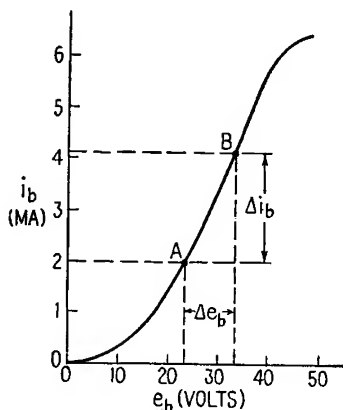


FIG. 3.3.—Diode i_b - e_b curve.

is maintained only for extremely short intervals, curves similar to those of Fig. 3.2 may be obtained.

Fig. 3.3 is the i_b - e_b curve for a diode operated at its rated cathode temperature. The *variational plate resistance* is defined as

$$r_p = \frac{\partial e_b}{\partial i_b} = \left. \frac{de_b}{di_b} \right]_{T \text{ const.}} = \lim_{\Delta i_b \rightarrow 0} \left. \frac{\Delta e_b}{\Delta i_b} \right]_{T \text{ const.}} \quad (3.2)$$

The increments Δi_b and Δe_b in Fig. 3.3 are measured from the point A. The *variational plate conductance* is the reciprocal of r_p , or

$$g_p = \frac{\partial i_b}{\partial e_b} = \frac{1}{r_p} \quad (3.3)$$

which is the slope of the i_b - e_b curve.

Consider a diode in the circuit of Fig. 3.4 in which

$$\bar{E}_{bb} = e_b + i_b R_L \quad (3.4)$$

where e_b is a function of i_b and R_L is the d-c resistance of the load.

If i_b is a known algebraic function of e_b , the equation expressing their relation can be solved simultaneously with (3.4). If, instead, the i_b - e_b curve for the diode (operated at rated cathode temperature) is available, a graphical solution can be obtained, Fig. 3.5, as follows:

By (3.4), $e_b = \bar{E}_{bb} - i_b R_L$. This is the equation of the line AB , Fig. 3.5, called the *d-c load line*. Though Ohm's law does not apply to the vacuum tube, it can be applied to the remainder of the circuit to determine AB , the locus of e_b for various values of current. For $i_b = 0$, $e_b = \bar{E}_{bb}$, locating the point B at the full battery potential, 250 volts, say. For an arbitrary value of current, say 2 ma, and $R_L = 50,000$ ohms, it can be seen that

$$e_b = \bar{E}_{bb} - i_b R_L = 250 - 100 = 150 \text{ volts,}$$

locating another point (such as P) on the line AB . The intersection of the load line AB and the i_b - e_b curve is the point Q (quiescent point), whose coordinates specify the actual values of current and voltage under quiescent conditions.

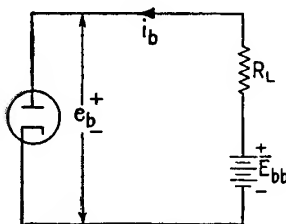


FIG. 3.4.—Diode in series with a resistor.

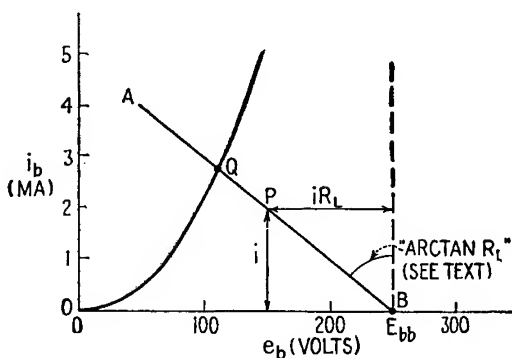


FIG. 3.5.—The d-c load line.

The construction of the d-c load line is most easily remembered by noting the position of the incremental voltage iR_L obtained for an arbitrary value of current, i . It is often stated that the load line is drawn at the angle $\arctan R_L$ with respect to a vertical axis. R_L can be considered to be the tangent of an angle in only a qualitative sense, since the increments of load voltage and load

current, rather than the distances that represent them, have the ratio R_L .

4. Space Charge.—The cloud of moving electrons between the electrodes of a vacuum tube constitutes a distribution called “space charge,” which is most dense in the immediate vicinity of the cathode where the electrons are moving at relatively low speeds. The forces on an electron in a diode include the force due to the potential applied to the plate and the force due to the space charge.

In a diode having parallel-plane electrodes, the force on an electron at various positions is as shown in Fig. 4.1. If the temperature of the cathode is such as to produce only a very low ther-

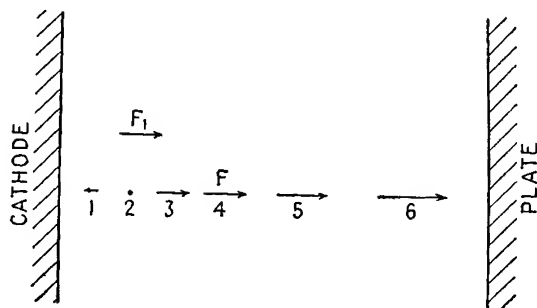


FIG. 4.1.—Force on an electron inside a diode.

mionic emission, the current and the space-charge effect are very small. Under these conditions the force on an electron at any position in the tube is F_1 , the force due to the uniform electric field set up by the plate potential. If, however, the thermionic emission is very high, so that the current is limited solely by space charge, the total force F on an electron is the resultant of F_1 and the space-charge force. Near the cathode, at position 1 in Fig. 4.1, the space-charge force is slightly greater than the force F_1 , so that the electrons leaving the cathode encounter a repelling force in the region between the cathode and position 2. At 2 the space-charge force just neutralizes F_1 . At position 3 the space-charge force is zero, and the net force on an electron is F_1 . To the right of 3, the space-charge force is directed toward the plate, so that the total force on the electron increases as the electron moves toward the plate (positions 5 and 6). During its motion beyond the equilibrium position 2, therefore, the electron has an *increasing* acceleration.

If the electrons leave the cathode with negligible initial velocities, the equilibrium position 2 for an electron is at the surface of the cathode. This means simply that the plate current increases until the space-charge force on an electron at the surface of the cathode is just sufficient to neutralize the force of attraction F_1 due to the plate potential. The actual initial velocities of the electrons cause the current to increase slightly beyond this value, shifting the equilibrium position 2 outward from the cathode, Fig. 4.1.

The theoretical space-charge-limited plate current in a diode is given by Child's formula

$$i_b = \frac{2.336 \cdot 10^{-6} e_b^{3/2}}{d^2} \text{ amp/cm}^2 \quad (4.1)$$

where e_b is the plate potential in volts and d is the distance in centimeters between cathode and plate. This equation was derived under the following assumptions:

1. The thermionic emission at the cathode is so great that the current is limited by space charge alone.
2. The electrodes are parallel-plane surfaces whose dimensions are large in comparison with the distance between them.
3. The electrons leave the cathode with negligible velocity.

The space-charge-limited plate current in a diode with *cylindrical* electrodes is

$$i_b = \frac{14.68 \cdot 10^{-6} e_b^{3/2}}{r\beta^2} \text{ amp/cm length of cathode} \quad (4.2)$$

Here β depends upon r/a , where r is the radius of the plate and a is the radius of the cathode.

$$\beta = \log \frac{r}{a} - \frac{2}{3} \left(\log \frac{r}{a} \right)^2 + \frac{1}{120} \left(\log \frac{r}{a} \right)^3 - \dots \quad (4.3)^1$$

5. Potential Distribution in a Plane Diode.—In Fig. 5.1 the potential at a point in a plane diode is plotted against distance from the cathode. At any point the electric field strength \mathcal{E} is given by the relation

$$\mathcal{E} = - \frac{de}{dx} \quad (5.1)$$

where e is the potential of any point with respect to the cathode and x is the distance from the cathode. The force experienced

¹ The derivation of (4.3) is described in "Fundamentals of Engineering Electronics," pp. 106–109, by W. G. Dow, John Wiley & Sons, Inc., 1937.

by an electron is proportional to de/dx , the slope of the potential curve.

The line (a) shows the distribution of potential when the space-charge effect is zero or negligible.

The curve (b) qualitatively describes the distribution of potential when the current is space-charge-limited and the initial velocities of emission of the electrons are negligible. The slope of this curve at the cathode is zero, indicating that the resultant field intensity at the cathode is zero.

The curve (c) shows qualitatively the actual distribution of potential, which differs from (b) because of the initial velocities of emission. Because of the initial velocities of the electrons,

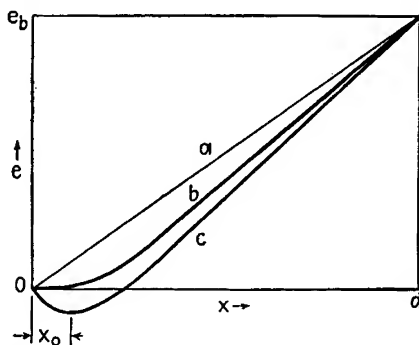


FIG. 5.1.—Potential distribution in a plane diode.

the current increases slightly beyond the value predicted by Child's law, (4.1), and the space-charge field at the cathode slightly exceeds the field produced by the plate. This means that the field in the immediate vicinity of the cathode, at a distance less than x_0 , is negative and that electrons emitted by the cathode experience a retarding force for a short distance before they reach the position x_0 of minimum potential. Beyond x_0 the electrons are accelerated toward the plate. Position x_0 corresponds to position 2, Fig. 4.1.

A mechanical analogy will serve to illustrate the ideas that have been expressed. If Fig. 5.1 is inverted, Fig. 5.2, and if the potential energy of an electron at the cathode is thought of as analogous to the potential energy of a marble at the top of an incline, the values of potential and field strength associated with various distances from the cathode can be interpreted in terms of height and the horizontal component of the force on the marble. As in Fig. 5.1, curve (a) shows the potential distribution for negligible

space charge, curve (b) for space-charge-limited current without initial electron velocities. Curve (c) is the actual potential distribution. An examination of curve (c) indicates that those electrons which are emitted by the cathode with sufficient velocity to get over the "potential hump" at x_0 will reach the plate, while those with lower velocities will return to the cathode. The elevation

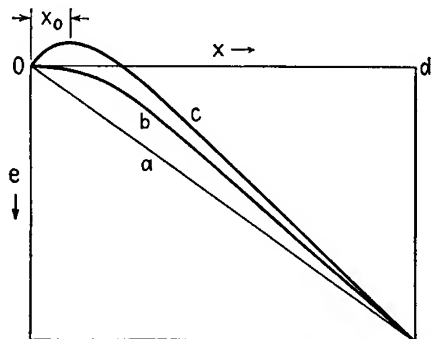


FIG. 5.2.—Potential diagram inverted.

of the potential hump and its distance from the cathode increase as the temperature of the cathode increases. If x_0 is thought of as the position of a "virtual cathode" from which the electrons are released without initial velocity and if the potential of the virtual cathode is taken to be that at x_0 , the assumptions underlying Child's law are reasonably justified. In Figs. 5.1 and 5.2 the distance x_0 and the region of negative potential are greatly exaggerated.

CHAPTER XI

MULTIELEMENT TUBES

1. The Control Grid.—The vacuum tube was a relatively unimportant device until the introduction of a control electrode, called the *grid*, forming a three-element tube, or *triode*. The addition of this control grid has led to a tremendous variety of applications of the vacuum tube. In most tubes, the grid is located between the cathode and the plate. Though it has commonly the form of a helix, it may be a mesh cylinder, a group of parallel wires connected together, or an electrode of some other shape.

When the current in a diode is space-charge-limited, the electrons themselves neutralize the field produced at the cathode by the positive potential applied to the plate. In fact, the current builds up until the space-charge force on an electron at the cathode slightly exceeds the attraction produced by the plate, so that the electrons leaving the cathode must move against a retarding force for a short distance, and only those having enough initial velocity will escape permanently from the cathode. Because the grid (of a triode) is located near the cathode, its potential is very effective in controlling the escape of electrons from the region near the cathode. A positive potential applied to the grid serves to reduce the retarding force on the electrons leaving the cathode, so that fewer return to the cathode, and the plate current is increased. A negative grid potential, on the other hand, increases the retarding force, thus causing more electrons to return to the cathode and decreasing the plate current. A sufficiently high negative potential (applied to the grid) will prevent the escape of any electrons from the region near the cathode, entirely cutting off the plate current.

Under the ideal conditions of Chap. X, Sec. 4, the space-charge-limited current in a plane diode is given by

$$i_b = \frac{2.336 \cdot 10^{-6} e_b^{\frac{3}{2}}}{d^{\frac{3}{2}}} \text{ amp/cm}^2$$

which can be written

$$i_b = \frac{2.336 \cdot 10^{-6}}{d^{\frac{3}{2}}} \left(\frac{e_b}{d} \right)^{\frac{3}{2}} \text{ amp/cm}^2 \quad (1.1)$$

Thus, for a given value of d , the current is proportional to the three-halves power of the space-charge-free field e_b/d at the cathode. This suggests that the effect of the grid upon the total cathode current in a *triode* is determined by its effect upon the space-charge-free field *at the cathode*. Let a quantity μ be defined, so that μe_c is the potential which, when applied to the plate, would produce the same effect upon the space-charge-free field as e_c does when applied to the grid. The space-charge-free field due to both plate and grid in an ideal triode is, therefore,

$$F_1 = \frac{e_b + \mu e_c}{d} \quad (1.2)$$

Since the (total) space-charge-limited current is proportional to the three-halves power of F_1 ,

$$i_b + i_c = A(e_b + \mu e_c)^{3/2} \quad (1.3)$$

where i_b is the plate current, i_c the grid current, e_b the plate potential, e_c the grid potential, and A a constant. Most low-power tubes are operated under conditions such that the grid current is negligible in comparison with the plate current, in which case

$$i_b = A(e_b + \mu e_c)^{3/2}$$

For an actual tube, the quantity μ is nearly constant throughout the most useful range of operating voltages. As in the diode, there is some departure from the three-halves-power behavior, the deviation being due in part to the initial velocity of electron emission, the effects of contact potential, and the geometry of the tube.

2. Characteristic Curves.—The properties of a three-element tube can be described most simply by graphs portraying the depend-

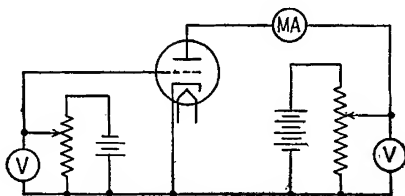


FIG. 2.1.—Circuit for studying triode characteristics.

ence of the plate current upon the values of grid potential and plate potential. Although a complete description of the properties of a tube must include also the dependence of grid current upon grid and plate potentials, the cases in which the grid current is negligible are of principal importance in low-power applications.

The circuit diagram of a laboratory setup for studying the characteristics of a triode is shown in Fig. 2.1. With the cathode

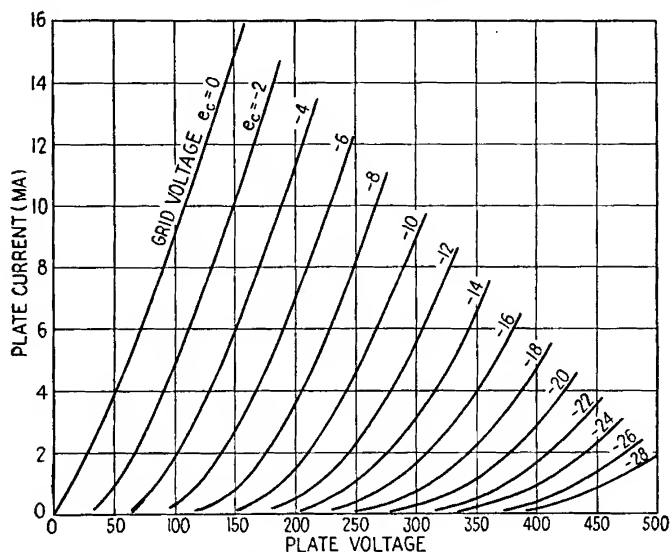


FIG. 2.2.—Constant grid-voltage (i_b-e_b) curves for a type 6C5 triode. (Data supplied by RCA.)

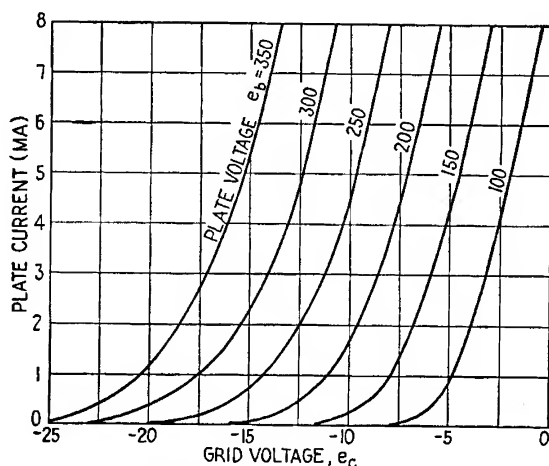


FIG. 2.3.—Constant plate-voltage (i_b-e_c) curves for a type 6C5 triode. (Data supplied by RCA.)

at the proper operating temperature, values of plate current can be obtained for various pairs of values of the grid and plate potentials. If the plate current i_b is plotted as a function of plate

potential e_b for each of several fixed (negative) values of grid potential e_c , a family of curves such as is shown in Fig. 2.2 is obtained.

The data used in plotting the curves of Fig. 2.2 may be represented graphically also by the family of curves in Fig. 2.3, in which plate current is plotted as a function of grid voltage, for each of several fixed values of plate voltage. A third representation, Fig. 2.4, is obtained by plotting plate voltage as a function of grid voltage, for fixed values of plate current.

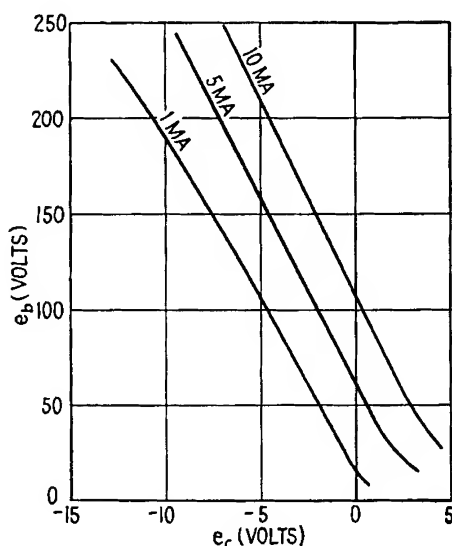


FIG. 2.4.—Constant plate-current (e_b - e_c) curves for a type 6C5 triode.

The simple circuit of Fig. 2.1 cannot be used to determine the tube characteristics when the steady currents are so great that the tube electrodes are overheated or the cathode is damaged. Special methods must be employed in which the high currents are drawn only during very brief intervals, so that neither overheating nor cathode damage occurs.

3. Operating Point, Path of Operation.—The *operating point* of a vacuum tube is the point (on a family of characteristic curves) whose coordinates represent the instantaneous values of the electrode potentials and currents. The *path of operation* is the path traversed by the operating point when a signal is applied. The path of operation can be represented on any of the three families

of curves, Figs. 2.2, 2.3, 2.4. The point of steady operation, without a signal, is called the *Q* point (quiescent point).

Consider the circuit of Fig. 3.1, in which the plate circuit contains a resistance R_L . At any instant, the plate voltage is given by

$$e_b = \bar{E}_{bb} - i_b R_L \quad (3.1)$$

The graph of this equation, on the i_b - e_b chart, is called the *load line*, *ABC* in Fig. 3.2. It is most easily constructed by choosing $i_b = 0$, for which

$e_b = \bar{E}_{bb}$, and some other convenient value of i_b , in order to obtain two points through which the load line can be drawn.¹ The *Q* point is located at the intersection of the load line

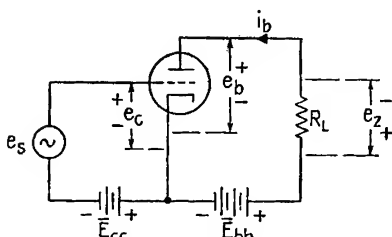


FIG. 3.1.—Triode with a resistive load in the plate circuit.

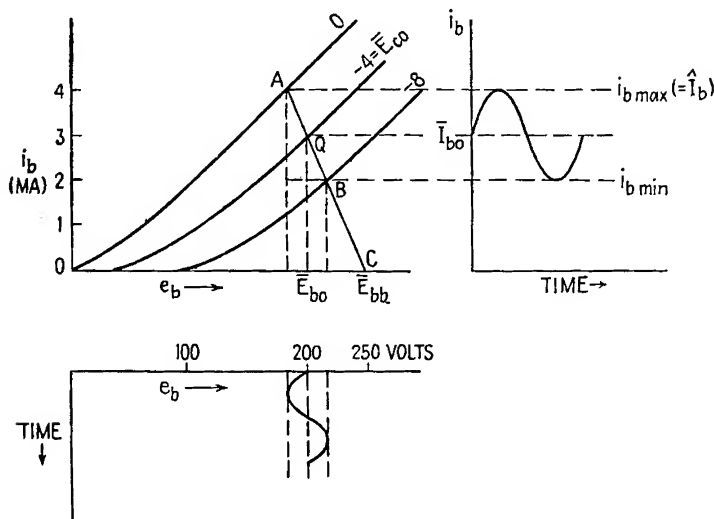


FIG. 3.2.—Path of operation on the i_b - e_b chart; $Z_L = \bar{R}_L = \bar{R}_L$.

with the i_b - e_b curve for the quiescent value of grid potential. In Fig. 3.2 it is assumed that in the absence of a signal the grid potential is -4 volts.

The quiescent values of plate current, plate voltage, and grid voltage are represented in Fig. 3.2 by the symbols \bar{i}_{b0} , \bar{E}_{b0} , and \bar{E}_{c0} .

¹ See Chap. X, p. 267.

As shown in Fig. 3.1, the signal e_s is introduced into the grid circuit in series with the grid-bias battery. As the grid potential is changed by the signal, the operating point moves back and forth along the path of operation AB . If the signal voltage is a sinusoidal function of time, the variations of grid potential, plate current, and plate voltage are as shown in Figs. 3.2 and 3.3. It should be observed that an *increase* in the grid potential (toward zero) is accompanied by an *increase* in plate current, but by a *decrease* in plate voltage (as a result of increased voltage drop in the load).

In many cases the a-c resistance of the load is different from its d-c resistance. Consider the circuit of Fig. 3.4, and assume the frequency of the signal high enough to make the reactance of C negligible in comparison with R_c . Then the a-c resistance of the load

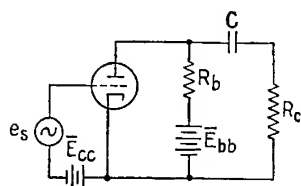


FIG. 3.4.—Vacuum-tube circuit in which the a-c load resistance is less than the d-c load resistance ($\tilde{R}_L < \bar{R}_L$).

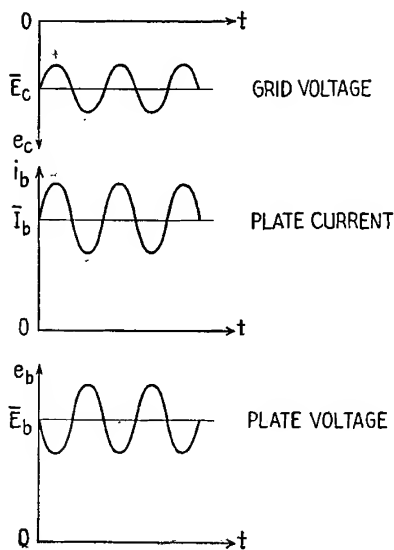


FIG. 3.3.—Grid voltage, plate current, and plate voltage for the path of operation of Fig. 3.2.

is that of R_b and R_c in parallel. The d-c load line is drawn for the d-c resistance R_b in the manner that has been described, and the Q point is located, after which the a-c load line can be constructed as AB in Fig. 3.5. Assuming linear operation, the a-c load line is drawn through the Q point at a slope such that

$$\Delta e_b = -\tilde{R}_L \Delta i_b \quad (3.2)$$

where \tilde{R}_L is the a-c resistance of the load, Δi_b is an arbitrary increment of plate current measured from the Q point, and Δe_b is the corresponding increment of plate potential. Equation (3.2) should be used in constructing the a-c load line. It is necessary merely to choose some increment of current Δi_b and multiply it by \tilde{R}_L to

maximum and minimum grid potential, as CD in Fig. 3.5. If the load is reactive as well as resistive ($Z_L = R_L + jX_L$), the path of operation is an ellipse, Fig. 3.6. The a-c load line intersects the ellipse at P and R , the points of maximum and minimum plate current. The operating point traverses the ellipse in a clockwise direction for an inductive load, in a counterclockwise direction for a capacitive load.

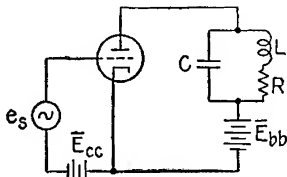


FIG. 3.7.—Tuned circuit as the plate load.

Example.—A circuit equivalent to that of Fig. 3.7 is often used as the load for a vacuum tube. Although its d-c resistance is usually very low, it may have a very high a-c resistance at the frequency of parallel resonance. The advantage of using such a circuit is illustrated in Fig. 3.8. The d-c load line is almost vertical, while the a-c load line AB and the path of operation CD at resonance are nearly horizontal. A very small signal can cause a large variation in plate voltage and, hence, a large a-c output voltage. If the signal is not at the frequency of resonance, the load will be reactive, producing an elliptical path

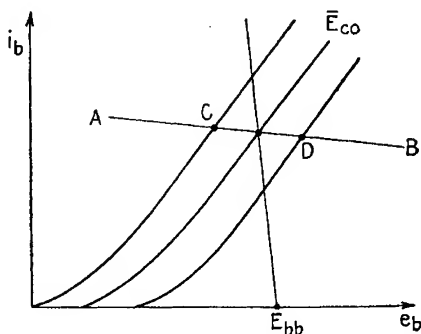


FIG. 3.8.—Path of operation for the load of Fig. 3.7.

of operation. This circuit is ordinarily used when only signals of frequencies near that of resonance are to be applied.

4. Power Relations in the Plate Circuit.—The power input to the plate circuit is equal to the power dissipated in the external circuit plus the power dissipated in the tube. In many important applications of the vacuum tube, a negative voltage and a relatively small a-c component of voltage (the signal) are applied simultaneously to the grid. If the average value of the plate current is not changed by the presence of the a-c component, as is the case when the signal is sinusoidal and the operation linear, the average power supplied by the battery will not be changed. Due to the

a-c component of current, however, there will be an increase in the average power supplied to a load resistor in the plate circuit, over the power due to the direct current alone. Since the power input is constant (when the d-c plate current is constant) and the power dissipated in the load increases, the power dissipated in the tube decreases. That is, the signal causes part of the power supplied by the battery to be transferred from the tube to the load.

If P_L represents the total average power delivered to the load, P_p the total average power dissipated at the plate of the tube, and P_{bb} the average power supplied by the battery,

$$P_{bb} = P_L + P_p \quad (4.1)$$

When P_{bb} is not changed by the signal and P_L is increased by the signal, P_p must decrease by a like amount.

5. Cathode Bias.—In Fig. 5.1 is shown an arrangement that

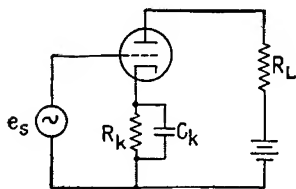


FIG. 5.1.—Circuit employing cathode bias, replacing C-battery bias.

provides proper bias voltage without the use of a C battery. Instead of employing a battery to keep the quiescent grid potential-negative with respect to ground, a *cathode resistor* R_k is introduced, whose voltage drop keeps the cathode at a potential that is positive with respect to ground. The grid is then placed at ground potential, in the absence of a signal, so that it is negative with

respect to cathode. The capacitor C_k is used to prevent variation of the voltage across R_k .

The graphical determination of the Q point for the circuit of Fig. 5.1 is illustrated in Fig. 5.2. The d-c load line is constructed in the usual manner except that the d-c resistance of the cathode resistor is added to the d-c resistance of the load to take account of the d-c voltage drop across the cathode resistor. The equation of the d-c load line now is

$$e_b = \bar{E}_{bb} - i_b(\bar{R}_L + \bar{R}_k) \quad (5.1)$$

The slope of the d-c load line is such that $\Delta e_b = -\Delta i_b(\bar{R}_L + \bar{R}_k)$. The d-c load line gives the value of the d-c *plate* voltage corresponding to any steady plate current.

The next step is to construct a line giving the value of the d-c *grid* voltage for any steady plate current. This line is called the

cathode-bias line. The potential of the grid with respect to the cathode is

$$e_c = -i_b \bar{R}_k \quad (5.2)$$

under static conditions. Hence, to determine points on the cathode-bias line, first choose convenient values of e_c . (The values of e_c corresponding to the curves on the i_b - e_b chart are most suitable.) Then determine the corresponding values of i_b from (5.2). These points are plotted on the i_b - e_b chart and are such points as O , a , b , c , d , Fig. 5.2. The line passing through these points is the cathode-bias line. It is not straight but is sometimes nearly so.

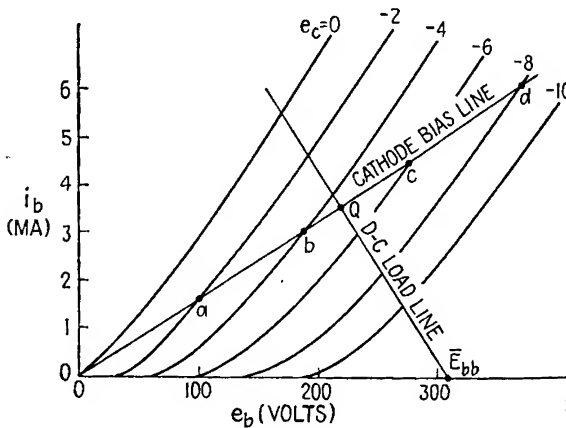


FIG. 5.2.— Q -point determination for a circuit employing cathode bias, Fig. 5.1.

The intersection of the cathode-bias line and the d-c load line is the Q point, Fig. 5.2.

The a-c load line, when required, is drawn through the Q point in the usual manner at a slope such that $\Delta e_b = -\bar{R}_L \Delta i_b$. The a-c load line is not influenced by the cathode-biasing resistor if it is adequately by-passed (by C_k) and the *average* value of the plate current does not change when a signal is applied.

6. Nonlinear Operation.—For the cases mentioned thus far, it has been assumed that the operating portion of the i_b - e_b chart consists of uniformly spaced parallel lines. Under these conditions the *quiescent* values of the various potentials and currents are the same as the *average* values that exist in the presence of the signal. When the lines on the i_b - e_b chart are not uniformly spaced, the relation between plate voltage and grid voltage is not linear. In the presence of a signal, therefore, the average values of plate

potential and current are not in general the same as the quiescent values. It becomes necessary to define an *A* point (average point) characterized by the average values of i_b and e_b , represented by the symbols \bar{I}_b and \bar{E}_b . Since the average plate current determines the d-c voltage drop in the load resistance, the *A* point must lie on the d-c load line. Also, the a-c load line must be drawn through the *A* point, as indicated in Fig. 6.1.

If the grid-bias voltage is independent of the plate current, the maximum and minimum values of grid potential will not be changed by nonlinear operation, in which case the a-c load line will be terminated at *M* and *N*, Fig. 6.1. If the grid bias depends upon the average value of plate current, as it does when a cathode

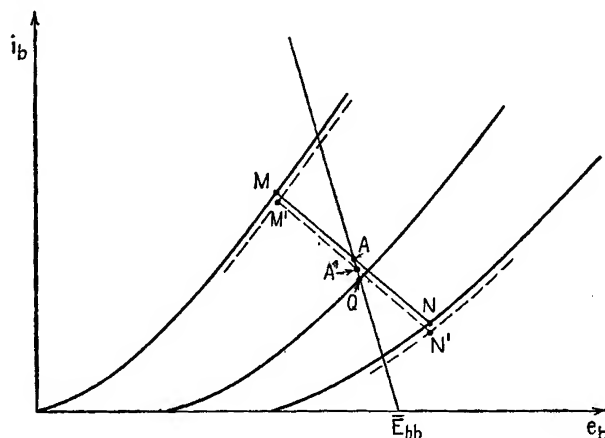


FIG. 6.1.—Nonlinear operation.

resistor is employed, the maximum and minimum values of grid potential will be changed to *M'* and *N'* and the *A* point will be at *A'*. It should be realized that, even under conditions of linear operation, the *A* point may differ from the *Q* point when there is a d-c component in the signal voltage applied to the grid.

7. Vacuum-tube Symbols.—In the study of vacuum-tube circuits, it is necessary to distinguish between direct and alternating components of current and voltage, as well as between the currents and voltages associated with the different tube electrodes and other parts of the circuits. For this reason, it is best to employ a somewhat extensive set of symbols, as listed in Appendix D and described below.

The instantaneous values of plate and grid current and voltage for a triode are represented by the symbols i_b (plate current),

e_b (plate voltage), i_c (grid current), e_c (grid voltage). The subscript indicates the position in the circuit (b for plate, c for grid, Fig. 3.1). The instantaneous voltage drop across the plate load is denoted by e_z . The d-c (average) components of current and voltage are represented by capital letters, accompanied by the proper subscripts, and are overscribed with a bar, viz., \bar{I}_b , \bar{E}_b , \bar{I}_c , \bar{E}_c , \bar{E}_z . The bar denotes average, or steady, values. The instantaneous alternating or variational components, measured from the average values, are denoted by small letters with the subscripts p for plate, g for grid, L for load, such as i_p , e_p , i_g , e_g , e_L . The rms values of these components are I_p , E_p , I_g , E_g , E_L , which ordinarily are considered as complex quantities. The quiescent values of i_b , e_c , etc., are indicated by the addition of the subscript 0 to the symbols used for

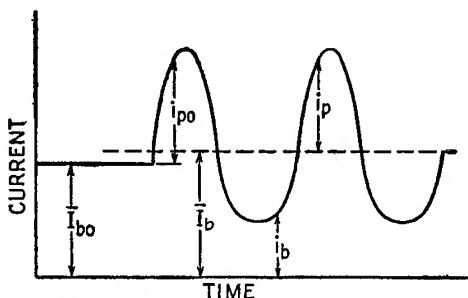


FIG. 7.1.—Illustrating symbols used in vacuum-tube notation.

average values: \bar{I}_{b0} , \bar{E}_{b0} , \bar{I}_{c0} , \bar{E}_{c0} , \bar{E}_{z0} ; and the variational components measured from the quiescent values are i_{p0} , e_{p0} , i_{g0} , e_{g0} , e_{L0} .

The symbols defined are illustrated for currents in Fig. 7.1, showing the plate current caused by an unsymmetrical signal, resulting in a difference between average and quiescent values. Nonlinear operation can also cause such a condition, even when the signal is sinusoidal. When the average and quiescent values are identical, the symbols for average values ordinarily are used.

8. Tube Coefficients.—If the plate voltage of a tube is held constant, a small increment Δe_c in the grid potential causes a small increment Δi_b in the plate current. The ratio $\Delta i_b / \Delta e_c$ may be thought of as a measure of the degree to which the grid voltage controls the plate current. The coefficient expressing this control is called the grid-to-plate *transconductance*, or *mutual conductance*, of the tube. Precisely, the transconductance is

$$g_m = \lim_{\Delta e_c \rightarrow 0} \left[\frac{\Delta i_b}{\Delta e_c} \right]_{e_b \text{ fixed}} = \frac{\partial i_b}{\partial e_c} \quad (8.1)$$

If the grid voltage is held constant, a small increment Δe_b in the plate voltage causes a small increment Δi_b in the plate current. The ratio $\Delta i_b/\Delta e_b$ may be thought of as a measure of the degree to which the plate voltage controls the plate current. The coefficient expressing this control is called the *plate conductance* of the tube. Precisely, the plate conductance is

$$g_p = \lim_{\Delta e_b \rightarrow 0} \left. \frac{\Delta i_b}{\Delta e_b} \right]_{e_c \text{ fixed}} = \frac{\partial i_b}{\partial e_b} \quad (8.2)$$

The reciprocal of g_p is the *plate resistance*, defined as

$$r_p = \lim_{\Delta i_b \rightarrow 0} \left. \frac{\Delta e_b}{\Delta i_b} \right]_{e_c \text{ fixed}} = \frac{\partial e_b}{\partial i_b} \quad (8.3)$$

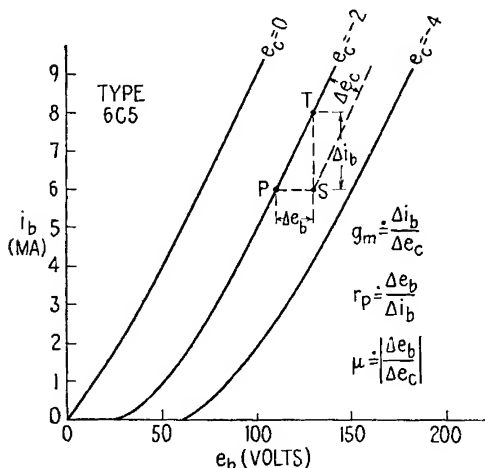


FIG. 8.1.—Illustrating tube coefficients.

The effect of an increment Δe_b of plate voltage on the plate current can be neutralized by an increment of opposite polarity in grid voltage. The ratio $|\Delta e_b/\Delta e_c|$ of such increments is a measure of the relative effectiveness of the grid as compared with the plate in controlling the plate current. The coefficient expressing this relative effectiveness is called the *voltage-amplification factor*, defined as

$$\mu = - \lim_{\Delta e_c \rightarrow 0} \left. \frac{\Delta e_b}{\Delta e_c} \right]_{i_b \text{ fixed}} = - \frac{\partial e_b}{\partial e_c} \quad (8.4)$$

The negative sign appears in (8.4) because the coefficient is conventionally defined as a positive number.

The coefficients g_m , r_p , μ are illustrated graphically in Fig. 8.1. Since g_m is the limit of the ratio $\Delta i_b / \Delta e_c$ as Δi_b approaches zero, it is proportional to the slope of the i_b - e_c curve at the operating point. A family of i_b - e_c curves is shown in Fig. 2.3. In Fig. 8.1, r_p is the limit of the ratio $\Delta e_b / \Delta i_b$ and is thus inversely proportional to the slope of the i_b - e_b curve. Also, $-\mu$ is the limit of $\Delta e_b / \Delta e_c$ and is proportional to the slope of the e_b - e_c curve, Fig. 2.4. Though μ , g_m , r_p can be evaluated from i_b - e_b or i_b - e_c or e_b - e_c curves, they can be measured much more accurately by bridge-type circuits designed for the purpose.¹

Typical values of triode coefficients and their dependence on plate current are shown in Fig. 8.2. It can be shown² that

$$\mu = g_m r_p \quad (8.5)$$

A graphical proof of (8.5) can be accomplished by means of Fig. 8.1. If, starting at P , the plate voltage is increased, causing a shift of the operating point from P to T , and then the grid voltage is decreased (made more negative), causing a shift from T to S , the two changes in plate current are

$$\Delta i_b \doteq \frac{\Delta e_b}{r_p} \quad \text{and} \quad \Delta i_b' \doteq g_m \Delta e_c \quad (8.6)$$

But these changes are equal in magnitude and opposite in direction, since the plate current at S is the same as the plate current at P . Therefore, $\Delta i_b' = -\Delta i_b$ and

$$g_m \Delta e_c \doteq -\frac{\Delta e_b}{r_p} \quad (8.7)$$

¹ E. L. CHAFFEE, "Theory of Thermionic Vacuum Tubes," p. 239, McGraw-Hill Book Company, Inc., 1933.

² Since $i_b = f(e_b, e_c)$,

$$di_b = \frac{\partial i_b}{\partial e_c} de_c + \frac{\partial i_b}{\partial e_b} de_b$$

If $di_b = 0$ (i_b fixed),

$$0 = \frac{\partial i_b}{\partial e_c} de_c + \frac{\partial i_b}{\partial e_b} de_b$$

from which

$$-\frac{de_b}{de_c} = \frac{\partial i_b}{\partial e_c} \frac{\partial e_b}{\partial i_b}$$

But

$$-\left. \frac{de_b}{de_c} \right]_{i_b \text{ fixed}} = \mu \quad \text{and} \quad \frac{\partial i_b}{\partial e_c} \div \frac{\partial e_b}{\partial i_b} = \frac{g_m}{g_p} = g_m r_p$$

Hence,

$$\mu = g_m r_p$$

Then

$$\frac{\Delta e_b}{\Delta e_c} \doteq -g_m r_p \quad (8.8)$$

But

$$\left. \frac{\Delta e_b}{\Delta e_c} \right]_{i_b \text{ fixed}} \doteq -\mu$$

Hence,

$$\mu = g_m r_p \quad (8.9)$$

Equation (8.9) is an exact relation, since all the approximate relations used in proving it are exact in the limit.

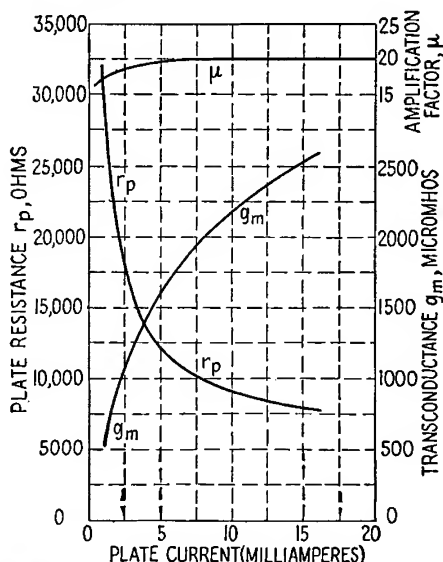


FIG. 8.2.—Showing μ , g_m , and r_p for a type 6C5 triode. (Data supplied by RCA.)

9. Equivalent Plate Circuit.—Because g_m and r_p are approximately constant over a restricted portion of the operating range of a vacuum tube, di_b , de_c , de_b , in (8.2), can be replaced by finite (but small) increments of plate current and grid and plate potentials. Thus

$$\Delta i_b = \frac{\partial i_b}{\partial e_c} \Delta e_c + \frac{\partial i_b}{\partial e_b} \Delta e_b \quad (9.1)$$

or

$$i_p \doteq g_m e_g + \frac{e_p}{r_p} \quad (9.2)$$

since the instantaneous values of *variational* plate current, etc., may be finite, small increments. For rms (complex) quantities,

(9.2) becomes

$$I_p = g_m E_g + \frac{E_p}{r_p} \quad (9.3)$$

showing that the a-c plate current can be expressed as the sum of two parts, proportional to the a-c grid voltage and the a-c plate voltage respectively.

Consider the circuit of Fig. 9.1. If there is an a-c component of plate current, the resulting a-c component of voltage in the load, E_L , will be accompanied by an equal a-c plate voltage of opposite sign. For since the sum of the load and plate voltages is a constant, \bar{E}_{bb} , any variation in one must be accompanied by an equal and opposite variation in the other. Then $E_p = -E_L$, and (9.3) becomes

$$I_p = g_m E_g - \frac{E_L}{r_p} \quad (9.4)$$

or

$$g_m r_p E_g = I_p r_p + E_L \quad (9.5)$$

But $E_L = I_p Z_L$, and $g_m r_p = \mu$; hence,

$$\mu E_g = I_p r_p + I_p Z_L \quad (9.6)$$

or

$$I_p = \frac{\mu E_g}{r_p + Z_L} \quad (9.7)$$

Equation (9.7) shows that the a-c component of current in the circuit of Fig. 9.1 is equivalent to the current in the circuit of Fig. 9.2, in which the load is connected to a generator whose no-load voltage is μE_g and whose internal impedance is r_p . Figure 9.2 usually is referred to as the *series or constant-voltage form of the equivalent plate circuit*. It is equivalent to the plate circuit insofar as alternating components of currents are concerned.

It is important to realize that μE_g is a fictitious, or equivalent, voltage. The only a-c voltages actually produced in the plate circuit are E_p and E_L , whose sum is zero. The voltage μE_g is

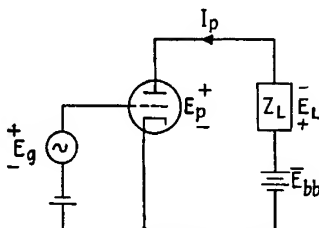


FIG. 9.1.—Plate circuit.

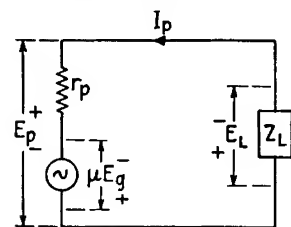


FIG. 9.2.—Equivalent plate circuit (constant-voltage form).

the voltage of a generator of internal impedance r_p , which will produce the same a-c load current as the vacuum tube with the signal applied to its grid. Actually, the a-c current is caused by the control action of the grid, not by an alternating emf in the plate circuit.

Example.—An alternating emf of 4 volts rms is applied to the grid of a 6C5, which is connected to a pure-resistance load of 100,000 ohms. Assuming that the tube is operated at its rated grid and plate polarizing potentials, the problem is to determine the a-c plate current and the voltage amplification.

The following data for the 6C5 tube are obtained from the RCA Receiving Tube Manual, Technical Series RC-14, page 100:

$$\begin{array}{ll} r_p = 10,000 \text{ ohms} & \bar{E}_{b0} = 250 \text{ volts} \\ g_m = 2,000 \text{ } \mu\text{mhos} & \bar{E}_{cc} = -8 \text{ volts} \\ \mu = 20 & \bar{I}_{b0} = 8 \text{ ma} \end{array}$$

From (9.7),

$$I_p = \frac{\mu E_g}{r_p + Z_L} = \frac{(20)(4)}{10,000 + 100,000} = 0.727 \text{ ma rms}$$

The *voltage amplification* is defined as the complex ratio of the a-c output voltage to the a-c input voltage, which in this case is the ratio of E_p to E_g . Since

$$E_p = -E_L = -I_p Z_L \quad (9.8)$$

from (9.7)

$$E_p = \frac{-\mu E_g Z_L}{r_p + Z_L} \quad (9.9)$$

Then the voltage amplification is

$$A = \frac{E_p}{E_g} = \frac{-\mu Z_L}{r_p + Z_L} \quad (9.10)$$

The minus sign indicates that, when Z_L is a resistance, there is a 180-degree phase difference between the a-c plate and grid voltages. Continuing the example,

$$A = \frac{-20 \cdot 100,000}{10,000 + 100,000} = -18.2 = 18.2/180^\circ$$

The magnitude of the voltage amplification is

$$|A| = \frac{|E_p|}{|E_g|} = \frac{|E_L|}{|E_g|} = \frac{|\mu Z_L|}{|r_p + Z_L|} \quad (9.11)$$

10. Phase Relations at Low Frequency.—In Figs. 10.1 to 10.3 are shown diagrams indicating the magnitude and phase relationships among the various alternating components of voltage and current for different types of load impedance in the circuit of Fig. 9.1.

For resistive load, Fig. 10.1, the output voltage E_p leads the input voltage E_g by 180° .

For inductive load, Fig. 10.2, the current I_p lags the voltage E_g , while E_L lags E_g . As a result, the output voltage E_p leads E_g by more than 180° .

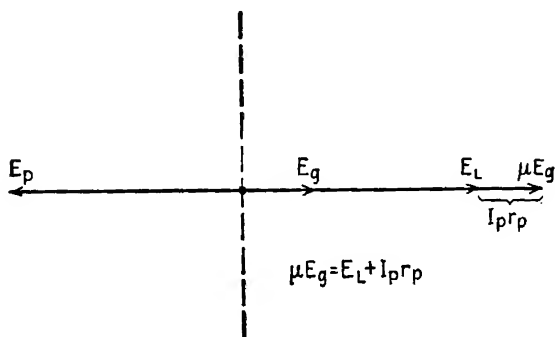


FIG. 10.1.—Phase relations for resistive load.

For capacitive load, Fig. 10.3, the current I_p leads the input a-c voltage E_g , while E_L lags E_g . The output voltage E_p leads E_g by less than 180° .

The diagrams shown are accurate only for low frequencies at which the reactances of the interelectrode capacitances of the tube are high in comparison with the other impedances of the circuit. At high frequencies, the effect of the interelectrode capaci-

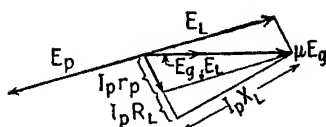


FIG. 10.2.—Phase relations for inductive load.

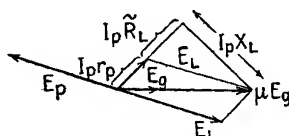


FIG. 10.3.—Phase relations for capacitive load.

ties of the tube is to cause the angle by which E_p leads E_g to be slightly smaller. As a result, the phase shift for resistive load is slightly less than 180° , and a certain amount of inductive reactance in the load is necessary to bring E_p into the 180° -degree position. This and other effects of the interelectrode capacitances are discussed later.

11. Constant-current Form of the Equivalent Plate Circuit.—

It is shown, (9.7), that

$$I_p = \frac{\mu E_g}{r_p + Z_L}$$

or $\mu E_g = I_p r_p + I_p Z_L$, (9.6). Substituting $g_m r_p$ for μ and dividing by r_p throughout,

$$g_m E_g = I_p + \frac{Z_L}{r_p} I_p \quad (11.1)$$

Consider the diagram of Fig. 11.1. If I_p is the current through Z_L , the voltage across Z_L , and therefore across r_p , is $I_p Z_L$, so that the current through the resistance r_p is $I_p Z_L / r_p$. Then the total current is $I_p + I_p Z_L / r_p$. This current is the same as that given by (11.1). Thus the vacuum tube can be represented by a source of constant a-c current $g_m E_g$, connected in parallel with r_p . The current through Z_L is, from (11.1),

$$I_p = \frac{g_m E_g}{1 + \frac{Z_L}{r_p}} = g_m E_g \frac{r_p}{r_p + Z_L} \quad (11.2)$$

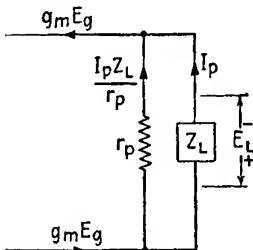


FIG. 11.1.—Equivalent plate circuit (constant-current form).

This equation and Fig. 11.1 describe the constant-current form of the equivalent plate circuit. When $r_p \gg Z_L$, a condition that often exists in circuits containing

pentodes, the constant-voltage and constant-current forms of the equivalent plate circuit yield the approximation

$$I_p \doteq g_m E_g \quad (11.3)$$

This approximate relation can be used with reasonable accuracy if, and only if, $r_p \gg Z_L$.

Example 1.—Consider a 6W7-G tube.

$$\begin{array}{ll} r_p = 1.5 \text{ megohms} & \bar{E}_{cc} = -3 \text{ volts} \\ g_m = 1,225 \text{ } \mu\text{mhos} & \bar{E}_{dd} = 100 \text{ volts} \\ \mu = 1,838 & \text{Let } Z_L = R_L = 10,000 \text{ ohms, and } E_g = 0.2 \text{ volt} \\ \bar{E}_{b0} = 250 \text{ volts} & \end{array}$$

Using the exact equation,

$$\begin{aligned} I_p &= \frac{\mu E_g}{r_p + Z_L} = \frac{1,838 \cdot 0.2}{1.5 \cdot 10^6 + 10^4} = \frac{3.676 \cdot 10^2}{1.51 \cdot 10^6}, \text{ or } 0.243 \text{ ma} \\ A &= \frac{-I_p Z_L}{E_g} = \frac{-0.243 \cdot 10^{-3} \cdot 10^4}{0.2} = -12.15 \end{aligned}$$

Using the approximation $r_p \gg Z_L$,

$$\begin{aligned} I_p &\doteq g_m E_g = 1,225 \cdot 10^{-6} \cdot 0.2, \text{ or } 0.245 \text{ ma} \\ A &= \frac{-I_p Z_L}{E_g} = \frac{-0.245 \cdot 10^{-3} \cdot 10^4}{0.2} = -12.25 \end{aligned}$$

Example 2.—As an example of a case in which the results obtained from the approximation (11.3) are not sufficiently accurate, consider a 6SF5 tube with a load impedance $Z_L = 5,000 + j0$. The tube constants are

$$r_p = 66,000 \text{ ohms} \quad g_m = 1,500 \text{ } \mu\text{mhos} \quad \mu = 99$$

Using the exact form (9.7), with $E_g = 0.2$ volt,

$$I_p = \frac{\mu E_g}{r_p + Z_L} = \frac{99 \cdot 0.2}{66,000 + 5,000}, \text{ or } 2.279 \text{ ma}$$

$$A = \frac{E_p}{E_g} = \frac{-I_p Z_L}{E_g} = \frac{-0.279 \cdot 10^{-3} \cdot 5,000}{0.2} = -6.97$$

Assuming $r_p \gg Z_L$ (which is not justified),

$$I_p \approx g_m E_g = 1,500 \cdot 10^{-6} \cdot 0.2, \text{ or } 0.3 \text{ ma}$$

$$A = \frac{-I_p Z_L}{E_g} = \frac{-0.3 \cdot 10^{-3} \cdot 5,000}{0.2} = -7.50$$

I_p and A are thus in error by 7.5 per cent when the approximate formulas are used, since $r_p = 13.2Z_L$ and is not in agreement with the condition $r_p \gg Z_L$.

12. Interelectrode Capacitances and Input Impedance.¹—The effects of the capacitances that exist between the electrodes of the tube were neglected in the treatment of the equivalent plate circuit in Secs. 9 and 10. In a triode there are three interelectrode capacitances: the capacitance between grid and cathode, C_{gk} ; the capacitance between grid and plate, C_{gp} ; the capacitance between plate and cathode, C_{pk} . These capacitances depend upon the physical dimensions of the tube; the larger the electrodes and the closer their spacing, the larger the capacitances. For small voltage-amplifier tubes these capacitances are usually less than 10 micromicrofarads ($\mu\mu\text{f}$). At low frequencies, they have very high reactances and can be neglected, but for frequencies higher than about 10 kilocycles per second (kcps) the effects of the interelectrode reactances cannot always be neglected.

Of particular interest is the effect of the interelectrode capacitances on the input impedance of a triode. The input impedance is defined as

$$Z_o = \frac{E_g}{I_g} \quad (12.1)$$

¹ F. L. SMITH, "Radiotron Designer's Handbook," 3d ed., pp. 46–48, RCA Manufacturing Co., 1941; W. L. EVERITT, "Communication Engineering," 2d ed., pp. 506–511, McGraw-Hill Book Company, Inc., 1937; R. S. GLASGOW, "Principles of Radio Engineering," pp. 228–235, McGraw-Hill Book Company, Inc., 1936; F. E. TERMAN, "Radio Engineering," 2d ed., pp. 231–234, McGraw-Hill Book Company, Inc.; H. J. REICH, "Theory and Applications of Electron Tubes," 1st ed., pp. 85–88, 2d ed., pp. 93–97, McGraw-Hill Book Company, Inc., 1939, 1944; CHAFFEE, *op. cit.*, pp. 261–281.

where E_g is the a-c component of the voltage applied to the input terminals (grid and cathode) and I_g is the alternating current entering the grid terminal. When the total grid potential is sufficiently negative, practically no electrons enter the grid and

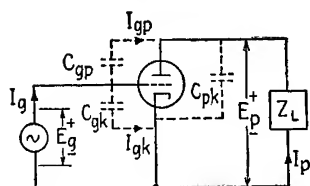


FIG. 12.1.—Internal capacitances of a triode.

the average, or d-c, component of the grid current is practically zero. When an a-c voltage is applied to the grid, the interelectrode capacitances carry an alternating current whose magnitude and phase depend upon the tube variational characteristics, and the plate-load impedance. Thus the input impedance

varies when the plate-load impedance is varied. This phenomenon is known as the Miller effect.

In the circuit of Fig. 12.1 (in which the polarizing sources are not shown), $I_g = I_{gk} + I_{gp}$, where I_{gk} is the alternating current through the grid-to-cathode capacitance and I_{gp} is the alternating current through the grid-to-plate capacitance. The a-c voltage across C_{gk} is E_g ; hence,

$$I_{gk} = \frac{E_g}{-j/\omega C_{gk}} = jE_g\omega C_{gk} \quad (12.2)$$

The a-c voltage from grid to plate is the vector, or complex, sum of the voltage from grid to cathode and the voltage from cathode to plate.

$$E_{gp} = E_g - E_p = E_g \left(1 - \frac{E_p}{E_g} \right) = E_g(1 - A)$$

The negative sign appears before E_p because E_p is the voltage from plate to cathode, which is the negative of the voltage from cathode to plate. Then

$$I_{gp} = \frac{E_{gp}}{-j/\omega C_{gp}} = jE_g\omega C_{gp}(1 - A) \quad (12.3)$$

and

$$I_g = I_{gk} + I_{gp} = j\omega E_g[C_{gk} + C_{gp}(1 - A)]$$

Hence,

$$Z_g = \frac{E_g}{I_g} = \frac{1}{j\omega[C_{gk} + C_{gp}(1 - A)]} \quad (12.4)$$

The input admittance (the reciprocal of the input impedance) is

$$Y_g = j\omega[C_{gk} + C_{gp}(1 - A)] \quad (12.5)$$

13. Input Impedance with Resistance Load.—A case of common occurrence is one in which the impedance Z_L is a pure resistance \tilde{R}_L . If the interelectrode reactances are large compared with \tilde{R}_L , the voltage amplification (9.10) is a negative real number, or $A = -|A| = |A|/180^\circ$. It should be understood that, unless the interelectrode reactances are very much larger than Z_L , the phase shift (for resistance load) is somewhat less than 180° . From (12.4),

$$Z_o = \frac{1}{j\omega[C_{gk} + C_{gp}(1 + |A|)]} \quad (13.1)$$

The input impedance is purely capacitive, and the value of the input capacitance C_i is

$$C_i = C_{gk} + C_{gp}(1 + |A|) \quad (13.2)$$

This capacitance can be many times greater than the value of C_{gk} or C_{gp} and increases as $|A|$ increases.

14. Input Impedance, General Case.—In general, the voltage amplification A is a complex number having a magnitude $|A|$ and a phase angle that we may denote by θ . Then

$$A = |A| \cos \theta + j|A| \sin \theta \quad (14.1)$$

Substituting this value in (12.5),

$$\begin{aligned} Y_o &= j\omega[C_{gk} + C_{gp}(1 - |A| \cos \theta - j|A| \sin \theta)] \\ &= |A|\omega C_{gp} \sin \theta + j\omega[C_{gk} + C_{gp}(1 - |A| \cos \theta)] \end{aligned} \quad (14.2)$$

Note that (14.2) contains a real term and a j term, signifying that the grid admittance consists of a conductance and susceptance in parallel.

For a resistive load, θ is 180° under the conditions of Sec. 13. If the load is reactive, the angle θ can change no more than 90° from the resistive case, so that

$$270 > \theta > 90^\circ$$

Then $\cos \theta = -|\cos \theta|$, and (14.2) may be written

$$Y_o = |A|\omega C_{gp} \sin \theta + j\omega[C_{gk} + C_{gp}(1 + |A| |\cos \theta|)] \quad (14.3)$$

Equation (14.3) indicates that the input susceptance is always capacitive, since the coefficient of $j\omega$ is always positive.

If the load reactance X_L is inductive within certain limits, θ is greater than 180° , $\sin \theta = -|\sin \theta|$, and the input conductance is negative. The derivation for the limits of X_L is rather lengthy

and can be found in various forms in the references listed at the beginning of Sec. 12. The results can be summarized as follows: If

$$\frac{\mu X_L}{\omega} > r_p R_L C_{gp} + (R_L^2 + X_L^2)(C_{gp} + \mu C_{gp} + \mu C_{pk}) \quad (14.4)$$

the input conductance is negative.

Another way of stating the above condition is that the input conductance is negative when X_L is positive and

$$X_L' < X_L < X_L'' \quad (14.5)$$

where X_L' and X_L'' are the roots of the equation obtained by equating the left-hand side to the right-hand side of (14.4). Writing

$$Z_o = \frac{1}{Y_o} = R_o + jX_o$$

from (14.2)

$$Z_o = \frac{|A|\omega C_{gp} \sin \theta - j\omega[C_{pk} + C_{gp}(1 - |A| \cos \theta)]}{(|A|\omega C_{gp} \sin \theta)^2 + \omega^2[C_{pk} + C_{gp}(1 - |A| \cos \theta)]^2}$$

When $\theta > 180^\circ$, R_o is negative.

When the input resistance or input conductance is negative, there is a component of the input current 180° out of phase with

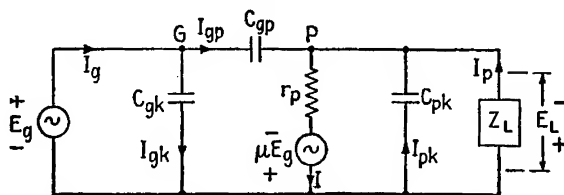


FIG. 14.1.—Complete equivalent circuit of a triode.

the voltage E_o . This means that power is being fed from the plate circuit to the grid driving circuit. If the grid driving circuit is a tuned LC circuit, the power fed back causes the voltage across the LC circuit to increase. This is called regeneration and may be used to increase the amplification at radio frequencies. If the power fed back is large enough, the circuit oscillates. This principle sometimes is used to produce oscillation. In an amplifier, oscillation is to be avoided, because in its presence the amplifier output is not completely under the control of the input signal.

The complete equivalent circuit of the triode, including inter-electrode capacitances, is shown in Fig. 14.1. The current and voltage relationships for this circuit are shown in Fig. 14.2. The

voltage μE_g is in phase with the voltage E_g . If the current I in Fig. 14.1 lags μE_g , as it will in most cases when Z_L is inductive, its phase is as indicated in Fig. 14.2. Then if $I r_p$ is subtracted from μE_g , the result is E_L , since, from Kirchhoff's laws, $\mu E_g = E_L + I r_p$. E_p is the negative of E_L , and $E_{gp} = E_g + E_L$. The current I_{gp} will lead E_{gp} by 90° . This current has a component 180° out of phase with E_g ; hence, power is being fed to the source of E_g . The input resistance and conductance are negative.

If the current I , Figs. 14.1 and 14.2, should lead μE_g , θ would be less than 180° and I_{gp} would have a component in phase with E_g . Consequently, E_g would be supplying power, and the driving source would be loaded by the tube. The input resistance and conductance would be positive.

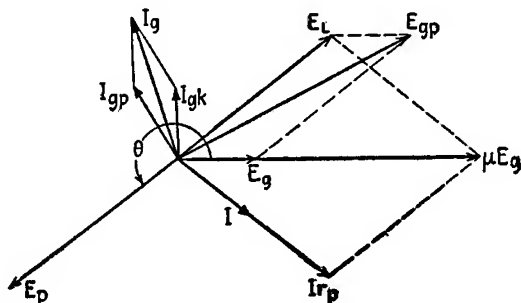


FIG. 14.2.—Vector diagram showing negative conductance.

I_{gk} and the quadrature component of I_{gk} always lead E_g , so that the input reactance and susceptance are always capacitive. If Z_L is inductive and has a high Q , the input conductance and resistance are negative for a wide range of values of Z_L . The sign of R_g for each of various types of loads is indicated in Table 14.1.

TABLE 14.1.—SIGN OF R_g FOR VARIOUS PLATE LOADS

Load	R_g
Pure resistance	Positive
Capacitive	Positive
Inductive	Negative for $X_L' > X_L > X_L''$; otherwise positive

15. Tetrodes and Pentodes.—The addition of a second grid (called a *screen grid*) between the control grid and plate of a triode forms a *tetrode*, or a four-element tube, Fig. 15.1. This grid is maintained at a steady positive voltage with respect to the cathode. It attracts electrons from the region surrounding the cathode, the electrons passing through the meshes of the screen grid to reach the

plate. Two functions are served by the screen grid. First, it reduces (by a factor of several hundred) the effective capacitance between control grid and plate, thus reducing the input admittance of the vacuum tube, by acting as an electrostatic screen or shield. This is of particular importance in high-frequency amplifier circuits, where the effects of high input admittance and coupling between input and output circuits are very troublesome. The second function of the screen grid is to reduce the degree to which the plate voltage influences the plate current (*i.e.*, to increase r_p without changing g_m appreciably), by reducing the effect of plate voltage on the space-charge-free field at the cathode. This effect

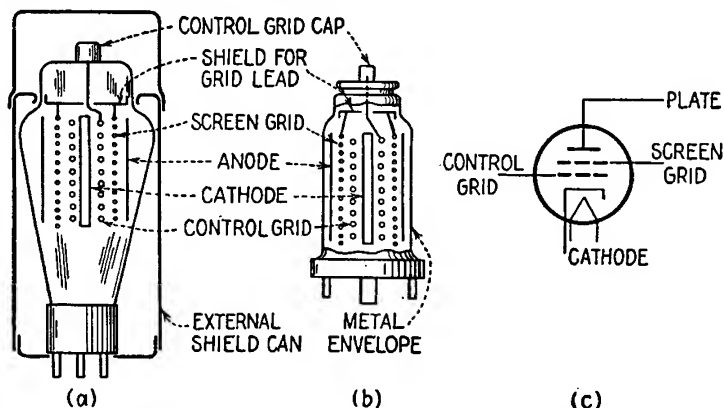


FIG. 15.1.—Screen-grid-tetrode, or four-element, tube: (a) construction employing a glass envelope; (b) construction employing a metal envelope; (c) tube symbol.

causes the plate current to be nearly constant over a large range of plate voltage. It also increases the a-c plate current for a given signal.

Consider the relation

$$I_p = g_m E_g + \frac{E_p}{r_p} \quad (15.1)$$

Since the a-c plate voltage is ordinarily approximately 180° out of phase with the grid voltage, the last term of the equation represents a reduction in the a-c plate current. If r_p is made very large, g_m remaining constant, the effect of the last term becomes quite small, so that the plate current is approximately $g_m E_g$. This reduction of the effect of the plate voltage is important in power amplifiers.

Because the screen grid does not greatly change the value of g_m from that of a triode having the same physical size, the value

of μ is increased by the same factor as r_p . Thus the tetrode is characterized by high values of both μ and r_p and by low grid-to-plate capacitance.

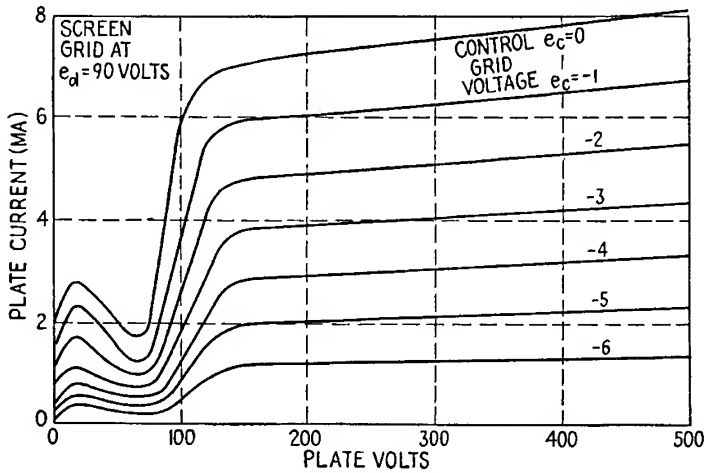


FIG. 15.2.—Typical i_b - e_b diagram for a screen-grid tetrode.

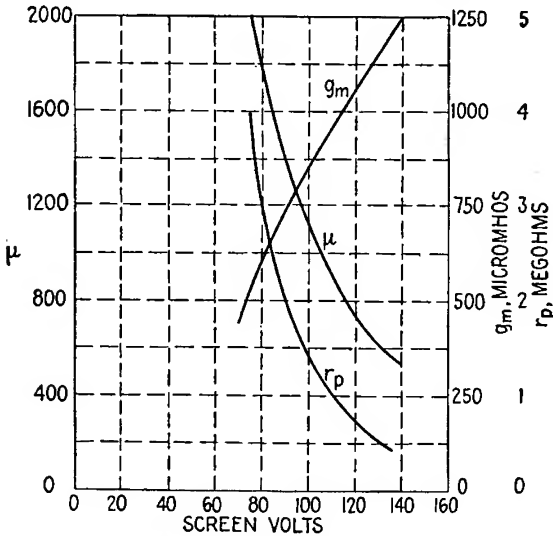


FIG. 15.3.—Effects of screen-grid voltage on μ , g_m , and r_p for a screen-grid tetrode

The i_b - e_b diagram for a 24-A screen-grid tetrode is shown in Fig. 15.2. The peculiar dips in the i_b - e_b curves in the region where the plate voltage is lower than that of the screen grid are the result

of the capture by the screen grid¹ of secondary electrons emitted at the plate. The effects of the screen-grid potential e_d upon the values of μ , g_m , r_p are shown in Fig. 15.3, in which the 24-A tube is represented as typical.

To obtain linear operation of the screen-grid tetrode, the path of operation must not extend into the region of the dips in plate current. The result is that the plate voltage cannot be allowed to

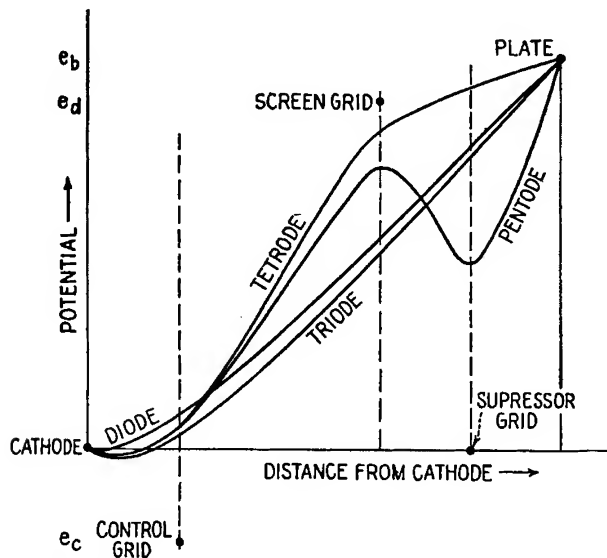


FIG. 15.4.—Variation of potential as a function of the distance from the cathode in a diode, tetrode, and pentode, along a path between grid wires. (The curves are accurate only in a qualitative sense.) Heavy dots indicate the distances and potentials of the grids.

decrease to very small values. For this reason, large amplitudes of a-c plate potential require a large d-c polarizing potential, thus increasing the power input to the plate circuit and the dissipation at the plate.

The undesirable effects of the capture of secondary electrons by the screen grid can be reduced by any one of several methods. The plate can be constructed of a material that does not readily emit secondary electrons; special fins or enclosures on the inner surface of the plate can be designed to reduce the number of escaping

¹ Figure 15.2 shows the characteristics of a tube whose plate is treated to reduce secondary emission. In an old-type 24-A tube the dips in the i_b - e_b curves are much more pronounced. In low-power applications the simple screen-grid tetrode is now obsolete.

secondary electrons. One of the most effective and widely used means of preventing the capture of secondary electrons by the screen grid is a "suppressor" grid placed between the screen grid and the plate. The addition of the suppressor grid makes the tube a pentode, or five-element tube. Kept at cathode potential, the suppressor grid establishes a low potential in the region between screen grid and plate, so that, in the space between suppressor grid and plate, electrons are acted upon by a large accelerating force toward the plate. Hence, secondary electrons leaving the plate soon lose their energy and are pulled back into the plate.

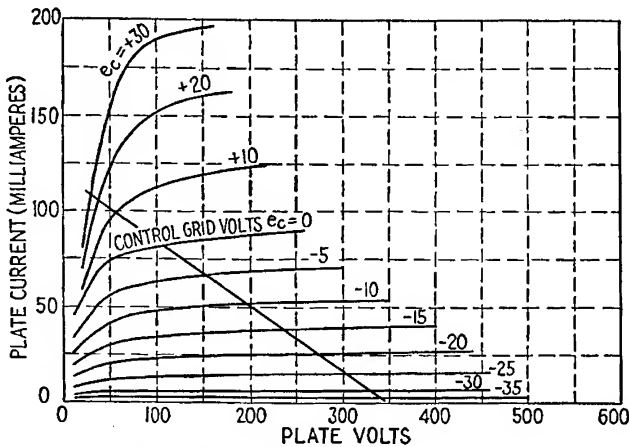


FIG. 15.5.—Typical i_b - e_b diagram for a power pentode.

Figure 15.4 compares the variations of potential (as a function of distance from the cathode) in a diode, a triode, a tetrode, and a pentode. The effects of the screen and suppressor grids are evident from the diagram, since the force on an electron is proportional to the slope of the potential curve.

The presence of the suppressor grid causes a decrease in the value of C_{sp} of a pentode in addition to that caused by the screen grid. Also, the values of μ and r_p are increased. There is no marked change in g_m . Typical values for μ , g_m , r_p for a small pentode such as the 6W7-G tube are $\mu = 1,838$, $g_m = 1,225 \cdot 10^{-6}$ mhos, $r_p = 1.5 \cdot 10^6$ ohms.

The i_b - e_b diagram for a pentode, Fig. 15.5, does not show the plate-current "dips" characteristic of the tetrode. For this reason the plate potential can swing to fairly low values without excessive nonlinearity. However, the load line, Fig. 15.5, shows

that, for linear operation, the path of operation cannot extend into the region of extremely low plate potential, because of the rounded "knees" of the i_b-e_b curves. This rounding off of the curves is one of the principal sources of distortion in the pentode, introducing a large third harmonic component of voltage into the plate circuit.

In the *beam-power* tetrode, both the suppressor-grid action and the elimination of the rounded knees of the pentode curves are attained through the use of "beam-forming" construction in the tube, illustrated in Fig. 15.6. The i_b-e_b diagram is shown

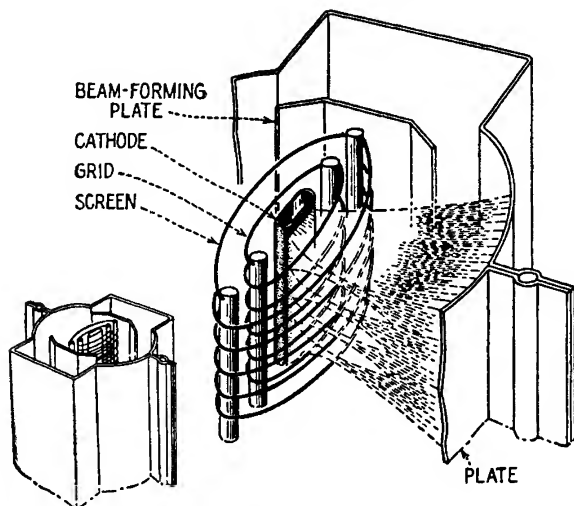


FIG. 15.6.—Internal structure of a beam-power tube (from *RCA Receiving Tube Manual*).

in Fig. 15.7. The beam-forming plates at cathode potential restrict the electron paths to beams at opposite sides of the cathode. The control- and screen-grid wires are aligned, with the screen-grid wires in the "shadow" of the control-grid wires, so that the electrons move toward the plate in concentrated sheets. The spacings of cathode, grid, screen grid, and plate are adjusted to critical values that produce the desired beam-forming effects. The net result is a concentration of the electron beams that produces a great space-charge effect in the region that would be the vicinity of the suppressor grid in a pentode. This concentration of space charge causes, and is in turn increased by, a great reduction in speed of the electrons as they approach the region mentioned, so that the effect of a suppressor grid is achieved. The improved characteristics of the i_b-e_b curves (sharp corners, improved linearity,

Fig. 15.7) result from the fact that the space-charge action which replaces that of the suppressor grid varies somewhat with the beam current. The beam-power tetrode can be operated with high efficiency because the screen dissipation is lower and a large ratio of a-c plate potential to d-c plate potential is attainable without undue distortion. The third harmonic distortion, one of the limiting factors in the operation of pentodes, is relatively small for a beam-power tube. The second harmonic distortion, however, is appreciable.

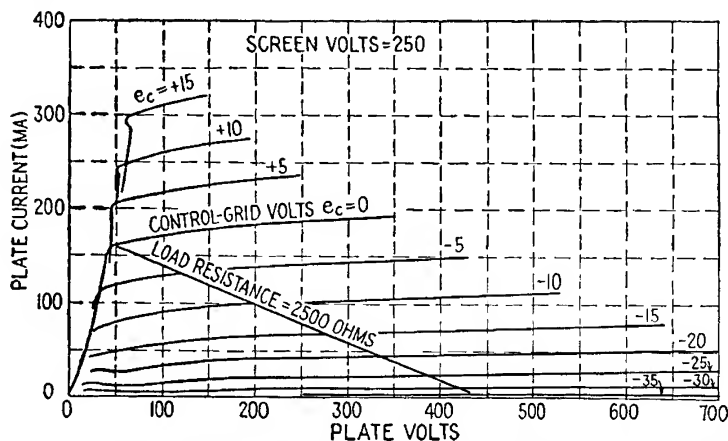


FIG. 15.7.—An i_b - e_b diagram for a beam-power tube, type 6L6 (from RCA Receiving Tube Manual).

16. Special and Multipurpose Tubes.—In Fig. 16.1b an i_b - e_c curve for a sharp-cutoff-type tube is contrasted with that of the remote-cutoff, or variable- g_m , tube. The remote-cutoff characteristic is obtained commonly by spacing the control-grid wires more widely near the center than at the ends, as in Fig. 16.1a.

As the grid potential is decreased from zero, becoming more and more negative, the electron stream first “cuts off” in the region near the ends of the spiral, where the spiral is tightly wound. Higher negative grid potentials confine the current to the center, or loosely wound, portion of the grid, where a complete and very gradual cutoff is achieved finally by a large negative grid potential. Near cutoff the tube has a low g_m . For grid potentials near zero, the remote-cutoff tube has a g_m which is intermediate between that for a tube with a tightly wound grid spiral and that for one of wider spacing.

In radio receivers the remote-cutoff tube is used as a high-

frequency amplifier. Since for a given load impedance the voltage amplification increases with g_m , the amplification obtained with the remote-cutoff tube can be controlled by adjusting the grid-bias voltage. The amplification is reduced in the presence of a strong r-f signal, so that the output of the receiver is more or less uniform

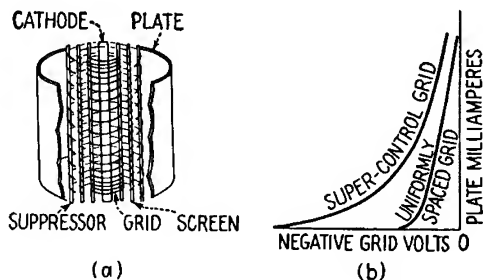


FIG. 16.1.—(a) Internal construction of a remote-cutoff tube; (b) comparison of its i_b - e_c characteristic with that of a sharp-cutoff tube (from RCA Receiving Tube Manual).

for all values of input signal above a certain minimum. This “automatic volume control” is more fully described in Chap. XXIII. The use of the remote-cutoff type of tube results in the reduction of cross modulation (cross talk) and certain other types of distortion that frequently arise in vacuum-tube circuits of high voltage amplification.

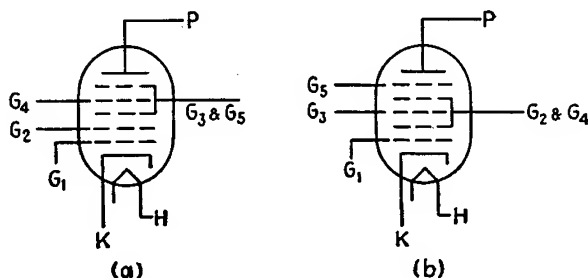


FIG. 16.2.—Symbols indicating the arrangement of electrodes in the order of their distance from the cathode for (a) a pentagrid-converter tube and (b) a hexodemixer tube.

Remote-cutoff tubes are called also “variable- μ ” tubes or “supercontrol” tubes.

For the normal, or sharp-cutoff, type of tetrode or pentode, the cutoff value of the control-grid potential is a function of the screen-grid potential e_a rather than of the plate potential e_p since

the screen grid shields the cathode from the effects of the plate voltage.

In Fig. 16.2 are shown the symbols for pentagrid-converter and -mixer tubes. These are special-purpose tubes used principally as frequency converters in superheterodyne radio receivers. Their use is explained in Chap. XXIII.

CHAPTER XII

CATHODE-RAY TUBES

1. Introduction.—A concentrated stream of electrons in an evacuated or partly evacuated space was described as a “cathode-ray beam” some time before the nature of the “beam” was clearly understood. Use of such a stream as a delicate indicator for electrical measurements is practically as old as the experimental study of the properties of such a beam. Electrons impinging on the glass walls of a vacuum tube with sufficient force indicate their arrival by the production of a faint glow. This is *fluorescence*, the excitation of visible light by a controlled disturbance of the molecules at the inner surface of the glass wall. These molecules are given extra energy, which they are able to retain for a brief period. They release this energy surplus in the form of visible light, its frequency or color depending upon the nature of the atoms and molecules struck by the electron stream.

In early experiments fluorescence soon was enhanced by painting the interior wall of the glass vessel with various chemical deposits. A particularly effective substance was the mineral *willemite*, which gives a greenish glow considerably stronger than that obtained from untreated glass. The fluorescence of willemite is due chiefly to zinc-orthosilicate, a synthetic substance now often employed. However, a very wide range of other colors and chemical substances is available. In fact, the millions of dollars of research spent in developing television receiving tubes led, as an almost incidental by-product, to the widespread development of fluorescent lighting. Essentially the cathode-ray tube is a source of fluorescent light in which the glow is produced by an intense concentrated stream of electrons, which travels widely over the fluorescent surface.

The candle power of the glowing spot is nearly proportional to the actual electric power dissipated at the screen. The equation for the candle power of the source is

$$CP = Ai(V - V_0) \quad (1.1)$$

where A is a constant, dependent upon the choice of screen material and electric units, i is the beam current, V is the potential of the

last anode, and V_0 is a small threshold potential necessary in order to just excite fluorescence.

After the electron stream has passed by, the candle power of the bombarded area does not instantly drop to zero. Normally it drops about 90 per cent in $\frac{1}{60}$ sec and then trails off gradually. This persistence of fluorescence is often advantageous in reducing flicker. Special tubes may be purchased with shorter persistence screens or with very sluggish screens that glow for a considerable period after bombardment.

Early methods for concentrating the electron stream were as simple and direct as the pinhole camera. A cloud of electrons was drawn toward a high-voltage anode. Most of the electrons struck the anode and were captured, while a few passed through a tiny hole into a field-free region beyond. Their inertia kept them going. Definition of the beam was improved by using a small cathode source and sometimes by using a second pinhole further to define the trajectory. Experimentally it was found that a small amount of residual gas narrowed the beam, positive ions in the path of the beam pulling the electrons toward its core. This was a crude sort of focusing action due to a radial component of electric field. It was never very satisfactory as the gas pressure was too critical and could not be held constant. Moreover, the current in the electron stream was too weak, and the screen had to be viewed in semidarkness.

Modern cathode-ray tubes are greatly improved in this respect; for the permissible aperture in the anode is much larger, and a considerable part of the total emission of the cathode contributes directly to the electron stream. This is possible because improved focusing devices are now available. The beam is now a pencil rather than a fine hairline. With appreciable thickness at most places along the axis, it narrows down to a fine pencil point just as it strikes the fluorescent screen.

2. Focusing, Electric-lens Method.—There are, in general, two methods for securing this focusing action. Each of them has acceptable variations, too numerous to describe in detail. One method, called the electric-lens method, makes use of a cylindrically symmetric radial component of the electric field. Instead of depending upon the uncertain action of residual positive ions, the tube is evacuated as thoroughly as possible, and the radial field is produced entirely by a series of cylindrical and conical electrodes, which surround the beam. Any series of cylinders,

strung along a common axis and operated at different electric potentials, may be referred to as a compound electric lens, for it guides electron trajectories in much the same way that a set of ordinary glass lenses guides rays of light. In particular, a small aperture in a low-potential electrode, followed by a larger aperture in a higher potential electrode, produces an over-all converging effect. The converging effect may be adjusted by a minor change of the ratio of the voltages on these two electrodes, usually produced by varying the lower voltage.

A typical *electron gun*, used for projecting an electrically focused stream of electrons, usually has two, three, or four individual

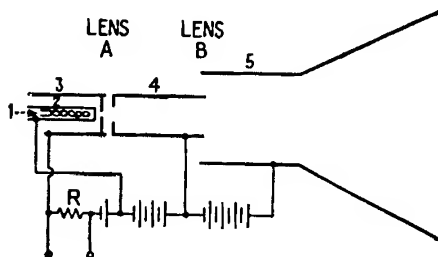


FIG. 2.1.—Simple lens system.

converging lenses of this character. The design problem is essentially similar to that of the modern electron microscope.

Figure 2.1 shows a particularly simple lens system of this character. Where batteries are indicated, equivalent "power-pack" voltages normally are employed. Lens *A* is formed by the adjacent apertures of electrode 3 and electrode 4. Lens *B* is formed by the similarly adjacent apertures of electrode 4 and electrode 5. Electrode 5 is often a carbon deposit held by a binder on a portion of the interior of the glass wall of the tube. Lens *A* is so strongly converging that the initially diverging electron stream is turned back toward the center and quickly crosses the axis, theoretically at a point. Actually, the "crossover" has a finite cross-sectional area but may be considerably smaller than the cathode from which the electrons started. The size of the crossover is of considerable importance, for the spot upon the fluorescent screen may be regarded as an image of the crossover, formed by the second lens *B*. Though *B* is also a converging lens, the converging action is not so strong. This time the electrons approach the axis gradually, meeting it at the screen, where a second crossover would occur if the screen were removed. In completely

empty space, after passing beyond the last anode, the electrons would gradually slow down and curve back toward the anode walls, where they would finally be captured. When the screen interrupts this journey, equilibrium is quickly established between the concentrated beam carrying the electrons toward the insulated obstacle and a diffuse cloud of electrons carrying an equal current from screen to anode. The action is similar to that of a powerful stream of water projected from a fire hose and striking a brick wall. The concentrated stream drives through a diffuse spray of returning water without being disturbed by the spray.

The alternative general method for producing a focused beam depends upon the production of an axial magnetic field. Consideration of this alternative method will be presented in Sec. 7.

3. Intensity Control.—A simple lens system for concentrating an electron beam having been described, the next consideration is *intensity control*. As indicated by Fig. 2.1, electrode 3 is placed in immediate proximity to the cathode and is operated at a potential moderately *below* that of the cathode. Active material such as barium oxide, which releases electrons at moderate temperatures, is deposited only upon the end wall of the cathode. From the outer portions of this circular area, electron emission is likely to be ineffective. Electrons, once released, are forced back into the cathode by the strong opposing field due to the negative electrode just ahead. From the central portion of this circular area, electrons may be urged forward by the accelerating field of the anode, effective through the aperture in electrode 3. One may think of a negative shadow, cast by this adjacent negative electrode upon the cathode surface, blighting the emission from areas sufficiently remote from the center. However, this negative shadow is determined not alone by the size of aperture and spacing of electrodes, but it depends in part upon just *how* negative electrode 3 is at the moment. The useful central area from which electrons come to join the axial stream may be increased or decreased by adjusting the steady-bias potential. By making this negative bias sufficiently large the electron stream may be cut off at its source. By applying an external alternating voltage, at a frequency of 1,000 cps for example, to the binding posts of resistor R , the effective area of the cathode is caused to fluctuate in size, modulating the electron stream and the spot of light that it produces. By analogy with the similar action exerted by the grid of the familiar triode amplifier tube, this control electrode is often referred to as a "grid." The name is justified by

the function, though not by the shape and construction of electrode 3.

Whenever the spot is stationary, the beam current should be held at the minimum usable value. Excess current directed at a single spot may "burn" the screen, producing a permanent discoloration and loss of fluorescence.

4. Deflection of the Beam by an Electric Field.—Suitable means for focusing and for controlling intensity having been found, it is now desirable to provide a method for deflecting the electron beam, vertically, horizontally, or by any combination of such deflections. The beam is a practically weightless pointer, which can move at speeds far beyond the reach of any mechanical pointer or moving mirror.

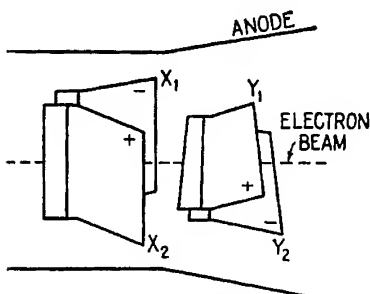


FIG. 4.1.—Deflecting plates.

Therein lies the greatest advantage of the instrument. Fundamentally, there are only two means by which a moving electron stream may be deflected without recourse to an actual mechanical collision. These are a transverse electric field and a transverse magnetic field. The *electric* method will be considered first, the magnetic method in Sec. 6.

Deflection by plates external to the envelope has had little application. The required voltage is too high, the operation is confused by charges induced upon the glass walls, and the focusing electrodes are at least partly in the way. If plates within the envelope are used, the sensitivity is improved by bringing the leading edges as near together as possible, without actual mechanical interference with the thick bundle of electron trajectories. The trailing edges must be spaced farther apart in order to allow room for gradual deflection of the rays as the electrons are pulled toward the positive plate and pushed away from the negative plate. This accounts for the V-shaped deflectors seen in most cathode-ray tubes. By applying a steady voltage to plates X_1 and X_2 , Fig. 4.1, the beam may be shifted to left or right of center. By applying a periodic voltage, the spot of light is caused to sweep back and forth along the horizontal or x axis of the screen.

To produce vertical deflection, a second pair of plates may be placed at right angles to the X plates. With plates close together there is no space for the X pair and the Y pair at the same location

on the axis. Hence they are displaced longitudinally, the electrons first experiencing a y deflection and then an x deflection, or vice versa. Aside from a minor difference in voltage sensitivity, the order makes no difference. As indicated in Fig. 4.1, both pairs of deflecting plates may be placed within the second anode, where the accelerating field is small and the electrons are moving with approximately constant forward velocity. At this position the transverse plates interfere very little with the lens action. In tubes having three or more anodes, the deflection is often introduced before the final acceleration. Before attaining their maximum forward velocity the electrons may be deviated more readily because they spend a longer time in the transverse electric field. This results in greater deflection sensitivity.

Linear amplifiers with constant gain over a specified frequency range usually are built into the oscillograph as conveniently available auxiliary devices. The deflection plates should also be directly accessible, however, for applications requiring a higher frequency range than is covered by the internal amplifiers.

An electron moving forward at constant speed in the presence of a uniform transverse electric field is deflected in a *parabolic* orbit, the direction of the acceleration of the negative charge being opposite to the field. Hence one may readily derive¹ the following expression for the *electric-deflection factor* of an idealized cathode-ray tube:

$$G_E = \frac{2d}{L} V \quad (4.1)$$

where G_E = deflection factor, in volts per unit distance of deflection upon the screen

d = spacing between parallel plates, expressed in same units

l = length of plates, expressed in same units

L = distance from center of plates to screen, expressed in same units

V = potential of the anode within which the deflecting plates are placed, measured with respect to the cathode

In practice, the deflection sensitivity of a practical 3-in. tube is of the order of 1,000 volts/m, usually expressed as 25 volts/in. By means of an appropriate amplifier contained within the oscilloscope box, the y -axis deflection factor is often reduced to approxi-

¹ J. MILLMAN and S. SEELY, "Electronics," p. 65, McGraw-Hill Book Company, Inc., 1941.

mately 0.1 volt/in., effective over the rated frequency range of the y amplifier. Similarly, the x -axis deflection factor (with a less expensive amplifier) is customarily reduced to approximately 0.7 volt/in.

Though (4.1) is useful in indicating the principal quantities that determine the *deflection sensitivity* (reciprocal of *deflection factor*) of a cathode-ray tube, it cannot be used in computing the actual deflection. In particular, it does not take account of the "fringing" of the field at the edges of the small capacitors formed by the deflection plates, nor does it allow for the V-shaped deflection plates now almost universally employed. Hence the actual deflection factor is measured experimentally.

The following remarks are general, in that they apply to particles of arbitrary mass and electric charge, shot through a transverse electric field that may vary in intensity at different positions along the z axis. For small deflections, such as are required in practice,

$$y \text{ deflection} = A_1 V_y \frac{e}{m} \frac{1}{v^2} \quad (4.2)$$

A_1 = proportionality constant dependent upon the units and dimensions

V_y = potential difference between y -deflection plates

e = electric charge carried by the moving electron or ion

m = corresponding mass of electron or ion

v = forward velocity, acquired before deflection (substantially uniform for all particles of a given type such as electrons, it will be very different for particles of some other type such as nitrogen ions)

Similar remarks apply to the x deflection. At first glance these general statements seem to imply that electrons and heavy ions will be deflected dissimilarly by an arbitrary transverse electric field. However, under the conditions encountered in all cathode-ray tubes, there is a further relationship among the quantities e , m , v . All the negative particles, whether heavy or light, have acquired their screenward velocity by falling through the same potential difference. Hence on equating potential energy lost by an arbitrary particle to the kinetic energy gained by the same particle,

$$\frac{1}{2}mv^2 = eV \quad (4.3)$$

whence

$$\frac{e}{m} \frac{1}{v^2} = \text{const. (independent of the type of particle)} \quad (4.4)$$

That is, when applied to a cathode-ray tube of the electric-deflection type, the more general expression of (4.2) reduces to

$$y \text{ deflection} \propto V_y \quad (4.5)$$

This means that the speed of the heavy ions is less than the speed of the electrons by an amount just sufficient to permit full deflection of the ions in spite of their much greater mass. Hence heavy negative ions will be deviated by deflection plates just as

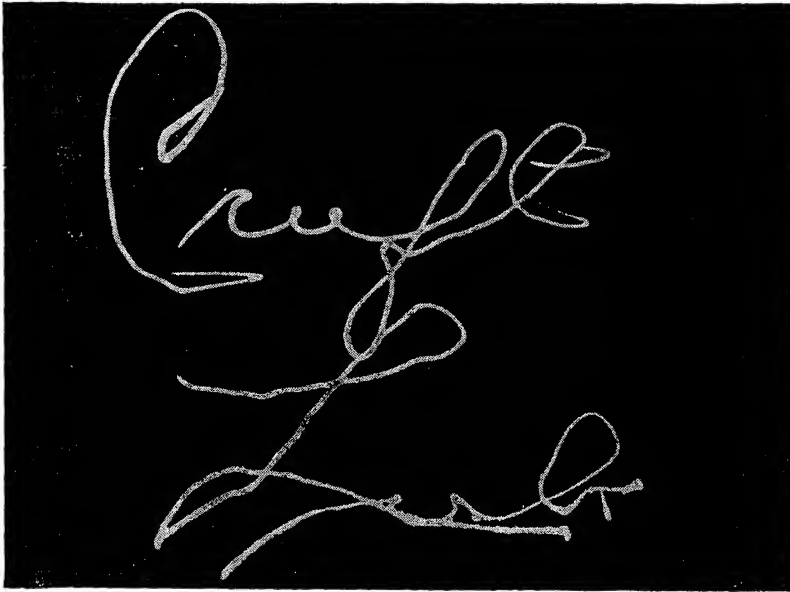


FIG. 5.1.—Pattern produced by x and y deflection and z -axis modulation.

much as electrons (and will also be brought to a focus by any electric lens designed to focus electrons).

5. Applications of Cathode-ray Tubes.—Common applications may be classified briefly as follows:

1. Plotting e_y vs. e_x .
2. Plotting i_y vs. e_x , or vice versa.
3. Plotting e_y vs. t_x or i_y vs. t_x .
4. Plotting e_y vs. f_x or i_y vs. f_x .
5. Plotting intensity vs. x and y at successive intervals of time (television).
6. Measuring short intervals of time (radar; ionospheric echoes; depth finding; submarine attack; numerous other applications).

Application 1 represents the most natural and direct function of the cathode-ray oscillograph. Since deflection originates in the electric field of two mutually perpendicular sets of deflection plates, a scale of voltage in both x and y directions is inherent in the cathode-ray tube. As a simple example one may demonstrate Lissajous figures, formed by connecting separate oscillators or other sinusoidal a-c generators to the x and y plates, Fig. 5.2. When

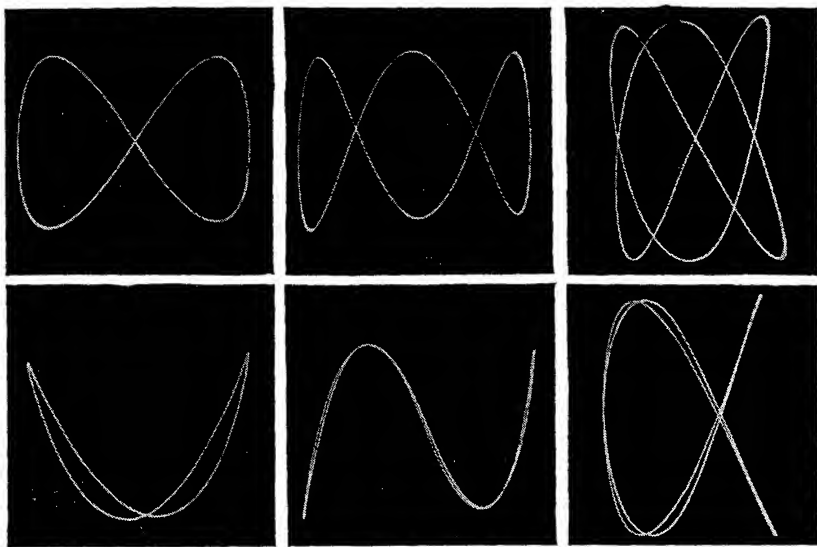


FIG. 5.2.—Lissajous figures.

the resultant closed figure becomes stationary, the frequencies of the generators are equal or are related by a simple ratio determinable from the shape of the steady pattern. This is useful in frequency calibration. When sinusoidal voltages e_y and e_x are identical in frequency, their phase angle may be determined from the shape of the resultant ellipse. Figure 5.1 illustrates the complexity of the cyclic pattern that may be formed by employing the proper waveforms for the x and y deflections.

Application 2 requires a “translating” device, capable of converting the inherent scale of voltage into a scale of current or of providing a current scale in addition to the voltage scale. A resistance connected across the binding posts for the x or y deflection serves this purpose, the translation being accomplished in accordance with Ohm's law. Examples of this application are

the instantaneous determination of vacuum-tube static characteristics, instantaneous examination of hysteresis loops, etc.

Application 3 is a particularly common one, needed whenever one wishes to examine or verify a waveform of current or voltage. For this "translation" one may employ a single linear sweep of voltage across the x axis, recording the y travel photographically, or may employ a recurrent saw-tooth wave applied to the x axis. In order that the observed y deflection shall retrace the same route on every successive sweep, facilitating visual observation, it is convenient to have means for varying the frequency of the saw-tooth wave. Preferably, the saw-tooth generator should be synchronized with the y -axis voltage wave under examination or should be synchronized at an exact submultiple frequency. The design of simple saw-tooth oscillators and their synchronization are discussed under Timing Circuits, Chap. XXIV. Such an oscillator is so common and useful as an auxiliary piece of apparatus that it is usually housed in the same cabinet with the oscilloscope, available for use when required. It is provided with a range adjustment, a fine adjustment of frequency, and a synchronization control.

Application 4 refers to the determination of resonance curves, modulation patterns, etc. A saw-tooth wave controls a frequency-modulated oscillator and at the same time sweeps the electron beam horizontally across the screen. The output of the device under test is displayed recurrently along the y axis as a function of the varying frequency. In "panoramic receivers" the amplitude of all signals received within a designated frequency band (say 200 kcps wide) is displayed as a function of frequency.

Application 5 is typified by television, though there are also important radar applications. Systematic changes of intensity paint in a picture, in lights and shadows, as the spot moves over a predetermined zigzag route, eventually covering all elements of area in a rectangular picture. Other types of systematic area coverage may be employed. The resultant area pattern is called a "raster." In television the entire display is renewed and repeated at a rapid rate, usually 30 times per second. The intensity control is referred to as a z -axis control to distinguish it from the x - and y -position controls. Z -axis amplifiers for intensity modulation are sometimes incorporated within the oscillograph cabinet.

Application 6 depends upon some form of calibration of the linear sweep voltages needed in applications 3 and 5. The time scale may be an external scale, pasted upon the tube face or engraved

on a transparent cover. More accurate results are obtainable with an *electronic* time scale, produced by jogging the beam or deflecting it sinusoidally, or by placing beads or gaps at regular intervals on the record. Any one of these time scales may be produced by means of voltages obtained from a timing oscillator whose frequency is accurately known or by means of a calibrated phase shifter. Often a circular sweep is provided. Special tubes may be provided with an axial electrode for convenient radial deflection.

6. Deflection of the Beam by a Magnetic Field.—A “magnetic” cathode-ray tube is one in which magnetic fields focus and deflect the beam. An electric field is always used for accelerating the electrons to produce the required forward velocity. Magnetic deflection possesses several advantages that apply particularly to tubes of large size. Tubes of 3- and 5-in. diameter found in simple laboratory oscilloscopes are usually of the “electric” type. Television tubes and other special-purpose tubes of 9-, 12-, or 20-in. size are often of the magnetic type.

The first advantage gained by magnetic focus and deflection has to do with the *shape* of the tube. The electron gun used in a magnetic tube is relatively simple and rugged and occupies little space along the axis. Especially in the larger sizes this results in a reduction of length in the stem, or “bottleneck,” of the tube. The entire glass structure is relatively shorter and stubbier and is less susceptible to damage from mechanical shock and vibration. Considering that space behind a panel is obtained often with great difficulty, the advantage is not so trivial as at first it may appear.

The second advantage of magnetic deflection is summarized by the expression, “Amperes are cheaper than volts.” Though not true in all engineering applications, this does apply here. Increased screen diameter requires higher candle power of the moving fluorescent spot, for the spot has less time to spend upon any particular square millimeter. To obtain this increased candle power both the beam current and the anode potentials are made larger. Higher electron velocity necessitates a more powerful deflecting force, no matter whether this be produced by electric or by magnetic means. If produced electrically, the voltage variation must have a greater amplitude and may require more expensive tubes in the deflection amplifiers and a power pack designed for higher voltage. For magnetic deflection, an external coil of wire is used with its axis at right angles to the shaft of the tube. Increased deflecting force requires an increase in the ampere turns of the coil, often obtainable by using inexpensive amplifier tubes at a larger fraction

of their current rating. Magnetic deflection in large tubes often leads to economy of weight, bulk, and cost.

However, magnetic deflection also has a disadvantage that is of major importance in some applications, but not in others. Unfortunately, the deflecting coil considered as a load to be driven by the deflection amplifier has *inductance* as well as resistance. In the "electric" tube the deviating electric field accurately follows all variation of the applied *voltage*. In the "magnetic" tube the deviating field depends upon the *current*, rather than upon the voltage delivered by the amplifier. When sinusoidal voltages are applied, the inductance produces a phase error. When any nonsinusoidal voltage is applied, the error is more serious, as the

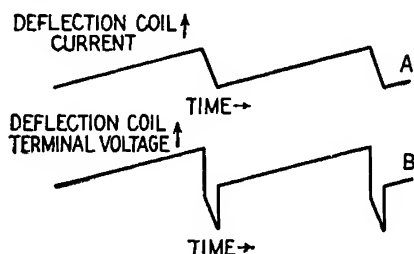


FIG. 6.1.—Deflection-coil current and terminal voltage as functions of time.

waveform of the current may be totally different from the waveform of the voltage. It is impractical to load the circuit so heavily with resistance that the distorting effect of the inductance will be negligible at the higher frequencies. Hence deflecting coils are unlikely to be used in a general-purpose oscillograph, such as a lecture-room instrument, even though a 20-in. screen is to be provided. On the other hand, if the deflection can be restricted to a very simple recurrent pattern, then provision can be made for the distortion effect of the inductance by applying a voltage having a special waveform. For example curve *B*, Fig. 6.1, presents the saw-tooth waveform that the magnetic field, and therefore the deflecting current, must have in order to produce a recurrent linear sweep. However, in accordance with the relation

$$e = L \frac{di}{dt} + Ri$$

the deflection amplifier must be driven by a voltage of peculiar waveform *B* (a "trapezoidal" voltage wave), in order that the current through an *LR* load shall follow the desired saw-tooth curve *A*. At first glance this may appear to offer an insurmountable

obstacle, even in the production of a simple linear sweep. Actually, the requisite trapezoidal wave may be generated and adjusted with sufficient accuracy and in a simple manner.¹ Though this is a special case, it is a particularly important special case, which includes television and a number of other area displays of the same general nature.

The deflecting coils may be mounted upon four iron poles, supported by a circular yoke. The poles project inward from the yoke, and their windings are connected in pairs to form crossed deflecting fields. The pole shoes are almost in contact with the outer surface of the glass stem of the tube. The iron structure is assembled from very thin laminations or from powdered iron held in shape by a cement "binder." The coils may also be used without iron. In this case the upper and lower halves of the vertical coil are wound in flat parallel spirals called "pancake coils." After winding, these flat coils are warped so as to follow the cylindrical contour of the glass stem and are then taped into a solid bundle of wire closely surrounding the stem. The halves of the horizontal coil may be included in the same bundle, overlapping the vertical coil.

The following discussion of the sensitivity of magnetic deflection intentionally parallels the corresponding discussion of electric deflection, thus emphasizing the similarities and the differences. An electron moving forward at constant speed in the presence of a uniform transverse magnetic flux is deflected in a *circular* orbit, the acceleration being perpendicular to the direction of the motion and the direction of the flux. Hence one may readily derive² the following expression for the *magnetic-deflection* factor of an idealized cathode-ray tube:

$$G_M = \frac{\sqrt{2}}{lL} \sqrt{V} \sqrt{\frac{m}{e}} \quad (6.1)$$

where G_M = deflection factor, in webers/square meter/meter

l = length of magnetic poles, in meters

L = distance from center of poles to screen, in meters

V = potential difference through which the charged particles have fallen before entering the deviating field

m = mass of particles, in kilograms

e = charge carried by particle, in coulombs

¹ Discussed in Chap. XXIV.

² MILLMAN and SEELY, *op. cit.*, p. 71.

Though (6.1) is useful in indicating the main factors that determine the *deflection sensitivity* of a cathode-ray tube (reciprocal of *deflection factor*), it cannot be used in computing the actual deflection. In particular it does not take account of the very extensive "fringing" of the magnetic flux at the edges of the widely spaced pole shoes. Hence the actual deflection factors are measured experimentally.

The following remarks are general in that they apply to particles of arbitrary mass and electric charge, shot through a transverse magnetic field that may vary in intensity at different positions along the z axis.

$$y \text{ deflection} = A_z \cdot B_x \cdot \frac{e}{m} \frac{1}{v} \quad (6.2)$$

where A_z = proportionality constant

B_x = magnetic flux density in the x direction

e = electric charge carried by the moving electron or ion

m = corresponding mass of electron or ion

v = forward velocity, acquired before deflection, substantially uniform for all particles of a given type

Similar remarks apply to the x deflection.

Qualitative inspection again indicates that the heavy particles will move slowly and that this low velocity will assist the deflection, therefore tending to offset the reduction of deflection occasioned by such a relatively large mass. However, in contrast to (4.2) the deflection now depends upon the inverse first power of the velocity rather than upon the inverse second power. Hence, on taking account of (4.3), one notes that the cancellation of terms is incomplete. When applied to an arbitrary cathode-ray tube of the magnetic-deflection type, the more general expression of (6.2) reduces to

$$y \text{ deflection} \propto B_x \sqrt{\frac{e}{m}}$$

Assume that an electric-lens system has brought a mixed beam composed of electrons and heavy ions to a single focus at the center of a fluorescent screen and that transverse magnetic flux is then established, resulting in a 5-cm deflection of the electron beam. Theoretically, the heavy-ion component of the beam is deflected in the same direction though by a different amount. However, upon remembering that the negative ions will have a mass thousands of times larger than the electronic mass, with little

or no increase of charge, it becomes evident that the 5-cm deflection of the electrons is accompanied by less than a millimeter deflection of the ion beam. For practical purposes, therefore, the negative ions are relatively unaffected by transverse magnetic flux correctly established for the deflection of the electrons. The same considerations show that a magnetic lens designed for focusing electrons will have relatively little converging effect upon negative ions.

7. Focusing, Magnetic-lens Method.—Magnetic focusing may be accomplished by establishing in an appropriate manner an *axial* component of magnetic flux. Ideally, one might consider a simple cathode-ray tube completely enclosed within a large single-layer solenoid, capable of producing a uniform axial magnetic flux throughout the *entire* space traversed by the electrons. On entering a long hollow anode the electrons would be accelerated quickly and then would retain a constant forward speed. In this simple case each electron would rotate about the axial flux lines while simultaneously moving forward toward the screen. The resultant motion would be a simple helical spiral. Though some spirals would have a large radius and others a small radius, the period of rotation would be identical.¹ Having been emitted from a point very near the axis of the tube, the electron would eventually return with a spiral motion to the immediate vicinity of the axis. By adjusting either the forward velocity or the rotational velocity, this return to the axis could be made to take place just as the screen would be struck, thus illuminating a small focal spot on the screen.

However, establishment of an adequately large flux density throughout such a large volume of space would be uneconomical and unnecessary. If an electron, diverging from the axis in a conical expanding beam, could suddenly enter a strong and uniform axial field, it would start to execute a tight spiral motion about the lines of magnetic flux. Having completed slightly more than 180° of the spiral travel, it should suddenly emerge from the limited space permeated by the axial flux. Meanwhile, the outward-diverging component of its motion necessarily would be exchanged for an inward component, thus causing the expanding cone of electron rays to be reshaped into a contracting cone, converging toward the desired focus. The focal distance could be adjusted by controlling the strength or the extent of the magnetic field. With practical engineering construction it would be difficult to establish very sharp boundaries for the origin and termination of

¹ Sec. 10, Appendix B.

the axial flux. The simple circular spiral motion is lost when a practical transition of flux density is permitted. Though this necessarily complicates the mathematical theory of the actual magnetic lens, it does not alter the over-all effect upon the electron stream—conversion of the expanding cone of rays into a contracting cone of rays—the essential change of course taking place in a limited portion of the z -axis travel.

In practice, therefore, magnetic focus is obtained by winding a multilayer coil around the stem of the tube. The axis of the focusing coil is approximately at right angles to each of the deflecting coils and coincides approximately with the axis of the tube. This arrangement forms a converging “magnetic lens,” which focuses the electrons at the screen. Such a magnetic lens may have “spherical aberration”—electrons traveling at different distances from the axis being guided to different foci. The lens designer reduces this effect by careful determination of the cross-sectional shape of the winding space, which ordinarily is not a simple rectangle. Provision is sometimes made for tilting the focusing coil slightly, so as to center the spot or deliberately to displace the spot from the center.

The advantages and disadvantages discussed in Sec. 6 apply chiefly to magnetic *deflection*. If one does not find it advisable to use magnetic deflection, then there is usually little advantage in using magnetic focus. On the other hand, if magnetic deflection does offer real advantages for the application in mind, then magnetic focus ordinarily will be specified in addition. Otherwise, the tube life may be shortened unduly by negative-ion bombardment of the center of the screen. This peculiarity may be explained as follows. Transverse electric fields treat electrons and heavy ions alike, Sec. 4, bringing them to the same focus and deflecting the two streams in exact coincidence. In the course of a day's work the stream of heavy ions, unavoidably present because of a trace of residual gas, has struck so many different portions of the screen that no one area is selectively damaged. The general deterioration of the screen is so slow that the tube may fail from other causes before the damage becomes noticeable. However, as shown in Sec. 6, magnetic flux has a strong effect upon electrons and a relatively weak effect upon heavy negative ions. As the available magnetic focus will fail to concentrate the heavy ions, they can cause no localized damage, regardless of the type of deflection employed. However, for the reasons already stated above, the combination of

electric focus with magnetic deflection is usually undesirable. Any electric lens, properly designed for focusing the electron stream, also causes the ion beam to be focused upon the center of the screen. This ion beam then remains in the central area, since the magnetic field that sweeps the electron beam is unable to produce any substantial deviation of the heavy ions. This cumulative destruction, concentrated daily upon the same small area, soon causes an objectionable discoloration. This is permissible only in special applications employing a circular sweep, where the central area will never be needed.

The magnetic lens and the electric lens are equally effective in producing a sharply focused beam in the small-size 1,000-volt tubes commonly employed as general laboratory instruments. For television applications and other services requiring larger tubes with higher accelerating potentials, the magnetic lens ordinarily has a slight advantage in retaining sharpness of focus.

8. Cathode-ray Oscillograph.—Figure 8.1 is a combination block and circuit diagram that shows basic features found in various common models of oscillograph. It does not correspond to any one model but represents a combination of parts of the Dumont models 164-E, 175-A, 208. Figure 8.2 is a complete circuit diagram of the Dumont Model 224-A.

The tube illustrated in Fig. 8.1 is a common type having a face 5 in. in diameter. All the connections to the electrodes are made through base pins; in the figure, leads from the first focusing anode and from the deflection plates are shown passing through the side of the tube in order to keep the figure simple.

The positive output terminal of one high-voltage power supply in the oscillograph is grounded. The negative terminal is attached to a potential divider to provide a variety of negative potentials, as indicated. The cathode (in the 164-E) is kept at $-1,000$ volts and the accelerating anode at 0 volts (chassis potential), so that the electrons move through a potential difference of $1,000$ volts in the process of acceleration. The intensity-control electrode may be made negative with respect to the cathode by any amount up to -100 volts. When it is made sufficiently negative, the beam is cut off completely. When the potential of the first focusing electrode is adjusted to a suitable value between -920 and -580 volts, the focusing fields have the correct strength to cause the electrons to converge toward a point on the screen.

Another high-voltage power supply, primarily intended to

supply plate voltage for the amplifiers and saw-tooth voltage generator, has its negative output terminal grounded. Its positive output terminal is connected to a potential divider so that (in the 164-E) a fixed potential of $+130$ volts exists at the point indicated in Fig. 8.1. Two 4-megohm potentiometers connected between the -130 - and $+130$ -volt points serve as positioning controls. For example, when the arm on the H-positioning (horizontal) control is to the right of the point of zero potential, an

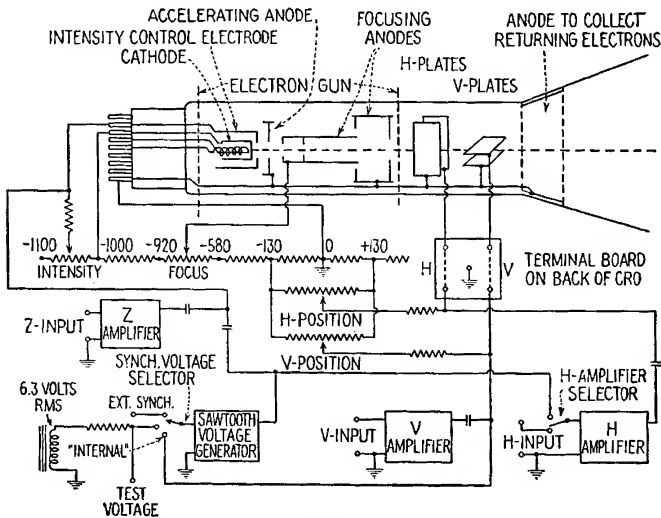


FIG. 8.1.—Block and circuit diagram of composite oscillograph.

electric field is created between the H plates that pulls the beam to the right (as seen by the observer). This positioning system is comparatively simple. (That in the 208 is more complicated and is not included here.)

The V-input (vertical) terminals are located in the lower left corner of the CRO panel. A signal voltage applied to the V-input terminals is coupled to the V plates either directly or through a one- or two-stage amplifier. Likewise, a signal voltage applied to the H-input terminals, located in the lower right corner of the panel, may be coupled to the H plates either directly or through a one- or two-stage amplifier. A terminal board is located on the rear of the oscillograph at the base of the tube. If desired, one or more of the links may be removed and a signal voltage applied to the H or V plates directly, thus eliminating the capacitance asso-

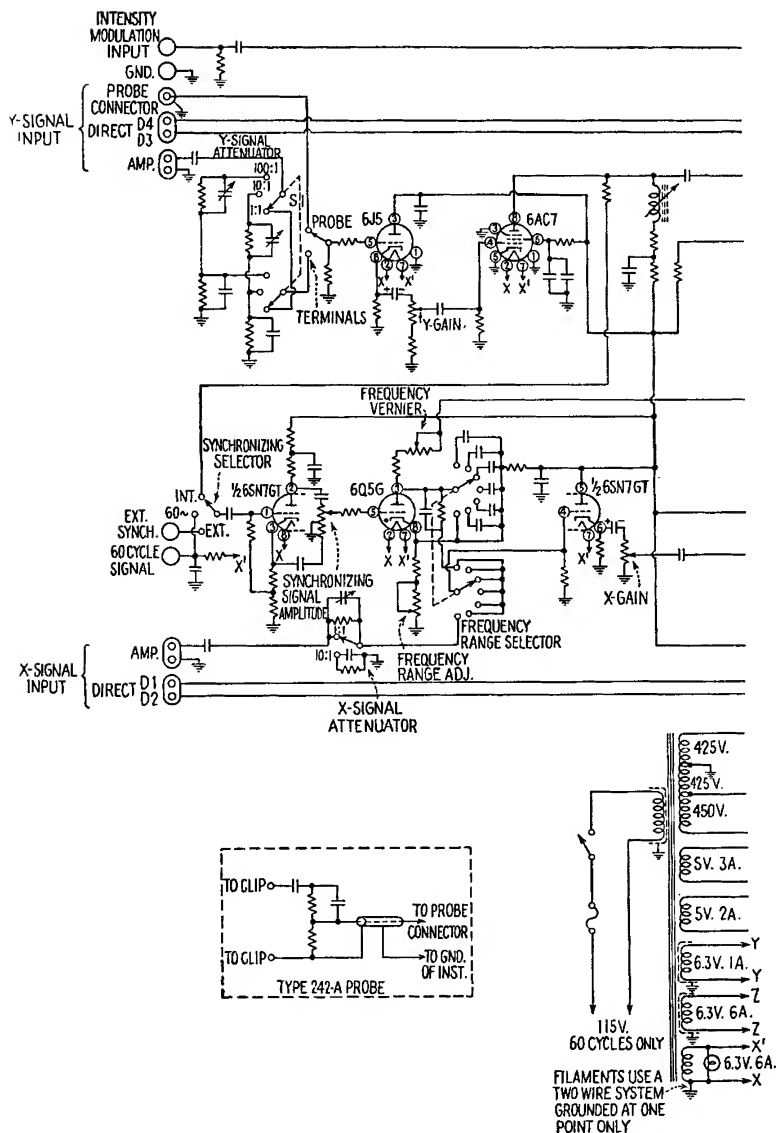
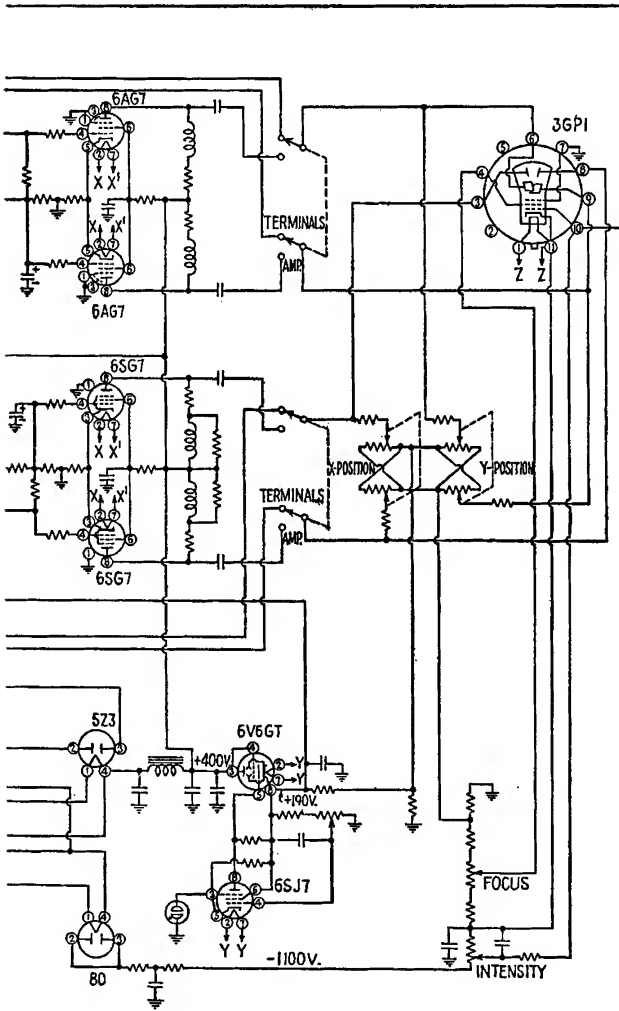


FIG. 8.2.—Schematic wiring diagram of the



Dumont-type 224-A cathode-ray oscillograph.

ciated with the wires inside the oscillograph that connect the front-panel input terminals to the terminal board on the rear.

When the saw-tooth voltage generator, Fig. 8.1, is connected to the H plates through the H amplifier by means of the H-amplifier selector switch, the saw-tooth voltage produces linear sweep action, *i.e.*, the spot repeatedly sweeps with uniform motion across the face of the tube and quickly jumps back to its starting point. The output of the saw-tooth voltage generator is also coupled, by means of a *CR* network, to the intensity-control electrode in such manner that a strong negative pulse is applied to the latter during the time that the spot is jumping back to its starting point. Thus, the spot is "blanked out" during its return trip. This operation is known as "return trace blanking."

The Dumont 224 and 241 oscillographs (but not the 164-E or the 208) are equipped with a Z input on the panel. A signal applied to the Z-input terminals is applied to the intensity-control electrode. The signal voltage thus introduced is superimposed upon and added to the steady direct voltage supplied by the intensity-control potentiometer, and the spot is blanked out every time the intensity-control electrode is suddenly swung in the negative direction by the Z signal. This operation is known as *z*-axis modulation.

The saw-tooth voltage generator may be synchronized with any one of several voltages. When the "synch" selector switch is in the "internal" position, the sweep frequency may be made equal to the signal frequency or a submultiple thereof, provided that the sweep-frequency controls are first set closely to the desired frequency. When the selector is in the "60 cycle" or "line voltage" position, the sweep circuit may be adjusted to 60 cps or a submultiple thereof, using the 60-cycle test-signal voltage supplied by a low-voltage secondary winding on the power transformer. When the selector is in the "ext-synch" position, the sweep circuit may be synchronized with a voltage applied to the "ext-synch" terminal and the "ground" terminal on the panel.

Figure 8.2 presents the complete schematic diagram of the circuit of a representative commercial model of cathode-ray oscillograph. Since discussions of amplifiers, attenuators, power supplies, and saw-tooth generators are presented elsewhere in this book, no detailed analysis of circuit components need be included here.

CHAPTER XIII

AMPLIFIERS—CLASS A AND CLASS B

1. The Vacuum Tube as an Amplifier.—The application of the amplifying properties of the vacuum tube has been largely responsible for the great advancements made in recent years in communications and in less well-known fields.

In a broad sense, the purpose of a vacuum-tube amplifier is to increase the voltage, or current, or power level of a signal to the value required to operate a given device. The device may require considerable power or it may depend for its operation upon high voltages with negligible power.

Voltage amplifiers are used to raise the voltage level of a given input signal. Low-power tubes with relatively high amplification factors are used. The degree to which a voltage amplifier raises the voltage of the input signal is indicated by its *voltage amplification*, which is the ratio of the output voltage to the input voltage. The complex voltage amplification is represented by A and its magnitude by $|A|$.

When the input signal E_s is applied to an amplifier of input resistance R_i and output resistance R_o , the voltage amplification is

$$A = \frac{E_o}{E_s} \quad (1.1)$$

where E_o is the output voltage. In general, E_o , E_s , and A are complex quantities. It is common practice to choose E_s as the reference voltage.

It is often advantageous to express the voltage amplification in terms of decibels. The gain G in decibels is defined as ten times the common logarithm of the ratio of the power output to the power input, or

$$\begin{aligned} G &= 10 \log_{10} \frac{|E_o|^2/R_o}{|E_s|^2/R_i} = 10 \log_{10} \left(\frac{|E_o|}{|E_s|} \right)^2 \frac{R_i}{R_o} \\ &= 20 \log_{10} \frac{|E_o|}{|E_s|} + 10 \log_{10} \frac{R_i}{R_o} \end{aligned} \quad (1.2)$$

If $R_i = R_o$, the gain of the amplifier is

$$G = 20 \log_{10} |A| \quad (1.3)$$

Both the over-all voltage amplification of a multistage amplifier and the voltage amplification of a single stage are of importance in most applications. In a multistage amplifier, the over-all voltage amplification is

$$A = A_1 A_2 A_3 \cdots \quad (1.4)$$

where A_1, A_2, A_3, \dots are the voltage amplifications of the individual stages.

Current amplifiers are in general of less importance than voltage amplifiers but have important applications when used for example, with photoelectric cells in control mechanisms. The *current amplification* of an amplifier has been defined in several ways. It is defined most logically, perhaps, as the ratio of the current delivered to the load when placed in the plate circuit of the amplifying tube to the current that would flow in the same load if it were connected directly to the source of emf.

Power amplifiers are used as output stages of audio-frequency amplifiers, output stages of transmitters, and driving stages in a transmitter and in any application where it is required that considerable power be delivered to the load. The load may be the input to a tube in which grid current flows for part of the cycle. The *power amplification* is the ratio of the alternating component of the output power to the alternating component of the input power to the amplifier. The power amplification often is expressed in decibels.

2. Classification of Amplifiers.—Amplifiers may be classified according to the frequency at which they operate, according to the band width to which they respond, according to the portion of the cycle of the input signal during which plate current flows, according to the nature of the network by means of which one stage is connected to the next.

Frequency Classification.—Amplifiers designed to amplify *direct currents or voltages* are called direct-current amplifiers. A direct current may be considered as having zero frequency. In general, direct-current amplifiers are effective also at relatively low frequencies of alternating voltages or currents. *Audio-frequency* amplifiers respond with varying degrees of uniformity to frequencies to which the human ear is sensitive. It is considered satisfactory in most high-fidelity systems if the response is restricted to frequencies from approximately 30 to about 15,000 cps. Indi-

vidual requirements cause wide variations in these limits, however. *Video-frequency* amplifiers are designed to amplify the video signal in television equipment. While the requirements for the frequency response of the video amplifier vary with the specific problem, it is not unusual to employ designs that give uniform frequency response in a continuous band from the lowest audio frequency (about 20 cps) to as high as 6 megacycles/second (mcps); the wider the band, the more detailed the picture. It is shown later that the upper frequency limit of such amplifiers is greater than 30 mcps. Amplifiers having the same general frequency response but used in other than television systems may be classified as *broad-band* amplifiers. *Intermediate-frequency* amplifiers are designed for the special purpose of amplifying the intermediate frequency of a superheterodyne receiver. The intermediate frequency varies from approximately 175 kilocycles/second (kcps) to many megacycles per second, depending upon the frequency band for which the receiver is designed. Contemporary practice in the design of broadcast receivers is to use an intermediate frequency of 456 kcps for amplitude-modulated transmission and 4.3 mcps for frequency-modulated broadcasts in the prewar 42- to 50-mc region. Higher intermediate frequencies are used in other applications. *Radio-frequency* amplifiers are used in amplifying radio-frequency signals. The radio frequencies extend from approximately 100,000 cps (long waves) to several thousand megacycles per second. The advent of television, frequency modulation, and other applications has stimulated research in the utilization of "ultra-high frequencies," or "microwaves."

Band-width Classification.—Audio-frequency and video amplifiers are designed ordinarily to respond to a frequency band that is large when compared with the mean frequency and are also classified as *wide-band* amplifiers. When the band of frequencies amplified is small compared with the mean frequency, the amplifier is a narrow-band (or tuned) amplifier. An example is the intermediate-frequency amplifier in an ordinary broadcast receiver, which has a band width of approximately 10,000 cps, whereas the mean frequency may be 456,000 cps. It is possible to have a narrow-band amplifier in the audio-frequency range. An example would be an amplifier to amplify 1,000 cps, with high attenuation for frequencies both below and above this value. The terms wide band and narrow band are merely descriptive; there is no sharp separation of the two classifications.

*Single-stage Classification.*¹—A single stage of an amplifier may be classified in accordance with the fraction of the cycle of the input signal during which plate current flows in the tube. These classes are: *Class A* amplifier, in which the grid-bias and signal voltages are such that plate current in a specific tube flows throughout the cycle; *Class AB* amplifier, in which the grid-bias and signal voltages are such that plate current in a specific tube flows for appreciably more than one-half cycle but less than the complete cycle; *Class B* amplifier, in which the grid-bias and signal voltages are such that plate current in a given tube flows for approximately one-half the cycle, whence it follows that the grid-bias voltage is approximately equal to the plate-current cutoff value and the quiescent value of plate current is approximately zero; *Class C* amplifier, in which the grid bias and signal voltages are such that plate current in a given tube flows for appreciably less than one-half cycle, whence the grid bias is appreciably greater than the plate-current cutoff value and the plate current is zero when no signal is applied to the grid.

Subscripts 1 and 2 are used to indicate the absence or presence of grid current. Thus, in a *Class AB₁* amplifier the grid never draws current; hence the grid never is driven positive during any part of the cycle. In a *Class AB₂* amplifier the grid may draw current during some part of the cycle, and the grid is driven positive during that part of the cycle for which grid current flows. *Class A* amplifiers are commonly operated as *Class A₁*. In the absence of the subscript, *Class A₁* operation may be assumed. *Class C* amplifiers are commonly operated as *Class C₂*. In the absence of the subscript, *Class C₂* operation may be assumed. The subscript should be used always when specifying *Class AB* and *Class B* operation.

3. Distortion in Amplifiers.—Ideally, the output voltage of an amplifier should have the same waveform as the input signal. Changes in the waveform are due to *distortion*. There are three types of distortion that may occur, either singly or together in any combination. They are nonlinear (amplitude) distortion, frequency distortion, and phase distortion. Frequency distortion and phase distortion have been defined in Chap. IX.

Nonlinear distortion exists when frequencies not present in the input signal occur in the output. The principal cause of nonlinear distortion is nonlinearity in the relation of either the grid

¹ "Standards on Electronics," Institute of Radio Engineers, 1938.

or plate current of a tube to the applied voltage. Nonlinear distortion is sometimes called *harmonic distortion*. When non-linearity exists, the relation between the current and voltage may be expressed in general by the series

$$i = \bar{I} + ae + be^2 + ce^3 + \dots \quad (3.1)$$

Assuming that the voltage e is given by

$$e = \hat{E} \sin \omega t \quad (3.2)$$

the current is

$$\begin{aligned} i &= \bar{I} + a\hat{E} \sin \omega t + b\hat{E}^2 \sin^2 \omega t + c\hat{E}^3 \sin^3 \omega t + \dots \\ &= \bar{I} + a\hat{E} \sin \omega t + \frac{b\hat{E}^2}{2} (1 - \cos 2\omega t) + \frac{c\hat{E}^3}{4} (3 \sin \omega t \\ &\quad - \sin 3\omega t) + \dots \\ &= \bar{I} + \frac{b\hat{E}^2}{2} + \left(a\hat{E} + \frac{3c\hat{E}^3}{4} \right) \sin \omega t - \frac{b\hat{E}^2}{2} \cos 2\omega t \\ &\quad - \frac{c\hat{E}^3}{4} \sin 3\omega t + \dots \quad (3.3) \end{aligned}$$

Thus the current contains a rectified component of magnitude $b\hat{E}^2/2$, the original frequency with amplitude $(a\hat{E} + 3c\hat{E}^3/4)$, the second harmonic with amplitude $b\hat{E}^2/2$, the third harmonic with amplitude $c\hat{E}^3/4$, etc. When the input signal contains more than one frequency, the number of new frequencies introduced becomes much greater, there being sum and difference frequencies in addition to the harmonics found here. Also, if higher power terms are required to represent the current-voltage relation, additional frequencies are present in the output.

Nonlinear distortion in an amplifier occurs also when grid current flows if either the internal impedance of the source or the external grid-circuit impedance is of appreciable value. When the impedance in the grid circuit is large, the voltage supplied to the grid of the tube is less than the emf of the source by the voltage drop in the external impedance. Since grid current flows only when the grid is positive, this decrease occurs during the positive half cycle of the applied voltage. The positive lobe of the alternating component of the plate current is distorted. This type of distortion is accentuated by the variation in the grid-bias voltage due to the rectifying action of the grid circuit.

Frequency distortion exists when all frequencies are not amplified by the same amount. It is due to the nature of the circuits asso-

ciated with the tube. If either the grid or the plate circuit incorporates reactive elements, frequency distortion occurs in some frequency range. The tube may cause frequency distortion if the period of the applied signal is of the same order of magnitude as the time of flight of the electron from the cathode to the plate of the tube.

A typical frequency-response curve for a transformer-coupled amplifier is compared with the ideal for a wide-band amplifier in Fig. 3.1. This amplifier has frequency distortion at the low and high frequencies but responds uniformly to frequencies between

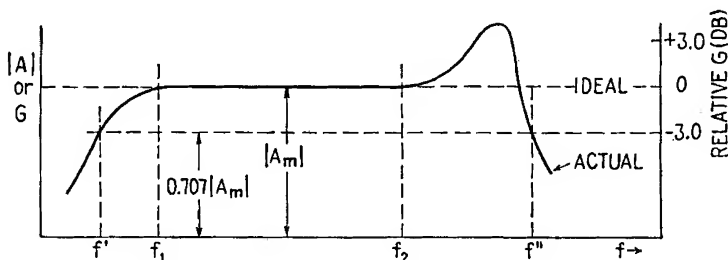


FIG. 3.1.—Typical frequency-response curve showing frequency distortion.

f_1 and f_2 . It is evident, therefore, that in the frequency range f_1 to f_2 this amplifier is free from frequency distortion.

The frequency response of an amplifier may be shown by plotting the relative gain in decibels as ordinate against frequency as abscissa. The reference gain usually is chosen in the mid-frequency range of the amplifier.

For the purposes of this treatment, the frequency-response curves are plotted with the magnitude of the voltage amplification as ordinate and the frequency as abscissa on a logarithmic scale.

The frequencies at which the power delivered to the load is one-half that delivered at some reference frequency, usually in the mid-frequency range, are important parameters in the study of amplifier characteristics. Since the power delivered to the load at a given frequency is proportional to the square of the voltage across it, the voltage amplification at the half-power frequencies is 0.707 times the voltage amplification at the reference frequency, or the gain of the amplifier is 3 db below the gain at the reference frequency. The lower and upper half-power frequencies f' and f'' are indicated in Fig. 3.1 for the amplifier having the frequency-response characteristics shown.

In multistage amplifiers, the half-power frequencies differ

from the half-power frequencies of the individual stages. For example, the upper half-power frequency of an amplifier that has two identical resistance-coupled stages is 0.64 of the upper half-power frequency of each stage, and the power delivered to the load at the upper half-power frequency of each stage is one-fourth the power delivered at the reference frequency. Similarly, in an amplifier consisting of identical stages, the voltage amplification of each stage at the half-power frequencies of the amplifier is given by

$$|A|' = 2^{-\frac{1}{2n}} |A_m| \quad (3.4)$$

where $|A_m|$ is the voltage amplification of each stage at the reference frequency. Thus, for a two-stage amplifier the half-power fre-

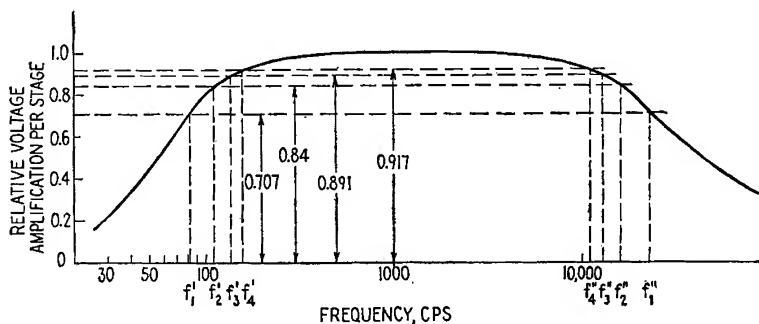


FIG. 3.2.—Typical frequency-response curve of one stage of amplification, showing the relative amplification per stage at the half-power frequencies of an amplifier having 1, 2, 3, and 4 amplifier stages (denoted by subscripts 1, 2, 3, 4).

quencies occur when the voltage amplification of each stage is 0.84 of the mid-frequency voltage amplification. The variation in the half-power frequencies as a function of the number of stages is shown in Fig. 3.2 for a typical wide-band amplifier.

Phase distortion occurs when the phase relations of the component frequencies of the output differ from those of the input signal. While phase distortion may occur without frequency distortion, the two usually exist simultaneously. Except at very high frequencies, phase distortion is introduced by the grid and plate circuits rather than by the tube.

Phase distortion is not important in audio systems because the human ear does not distinguish small changes in the relative phases of components of a sound. It is very important, however, when the amplifier is to be used with a cathode-ray tube for the purpose of studying waveforms. Phase distortion causes a change of wave-

form as shown in Fig. 3.3, where two waves, each consisting of a fundamental and a third harmonic, are compared when the third harmonic in the output has a relative phase shift of 180° . In a multistage amplifier the phase shift is given by

$$\beta = n\pi + \theta \quad (3.5)$$

where n is the number of stages and θ is called the *relative phase angle*. Furthermore,

$$\beta = \beta_1 + \beta_2 + \beta_3 + \cdots \quad (3.6)$$

where $\beta_1 = \pi + \theta_1$, $\beta_2 = \pi + \theta_2$, $\beta_3 = \pi + \theta_3$, \cdots are the phase

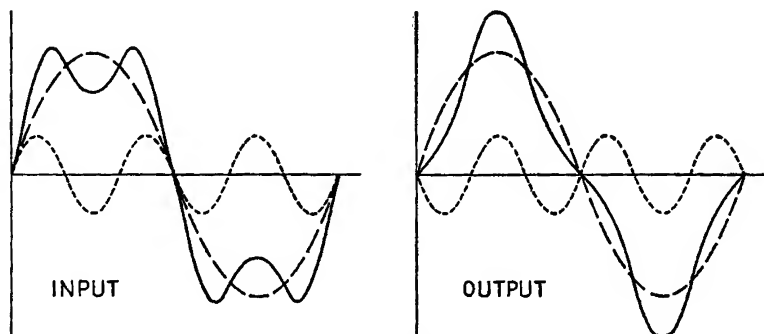


FIG. 3.3.—Effect of phase distortion.

shifts for the individual stages. The relative phase angle at a given frequency, therefore, is given by

$$\theta = \theta_1 + \theta_2 + \theta_3 + \cdots \quad (3.7)$$

Phase distortion is absent when θ either is zero or is proportional to the frequency of the component, *i.e.*, when the relative time delay θ/ω is constant.

4. Cascading of Amplifier Stages.—Seldom does a single-stage amplifier provide the voltage amplification required. It is necessary that many stages be connected in cascade when large amplification is desired. The type of network used to couple one stage to another is determined by the requirements of the amplifier. The various types of coupling commonly used are direct coupling, resistance coupling, impedance coupling, untuned-transformer coupling, tuned-transformer coupling.

5. Direct-coupled Amplifiers.—In direct-coupled amplifiers a d-c connection exists between the plate of the first tube and the

grid of the following tube. Only two of the many available connections are presented here. The circuit diagram of Fig. 5.1 shows one of the simplest types of direct coupling but by no means

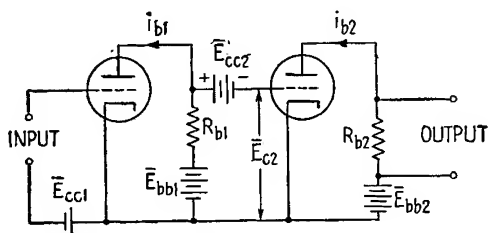


FIG. 5.1.—Direct-coupled amplifier.

the most desirable. In order to provide the proper grid-bias voltage for the second tube, the emf of the battery \bar{E}_{cc2} must be

$$\bar{E}_{cc2} = \bar{E}_{bb1} - \bar{I}_{b1}R_{b1} - \bar{E}_{c2} \quad (5.1)$$

This amplifier has the same amplification for d-c voltages and a-c voltages of frequencies up to those for which the reactance of the tube and wiring capacitances is of the same order of magnitude as R_{b1} . Small variations in the polarizing voltages may be of equal

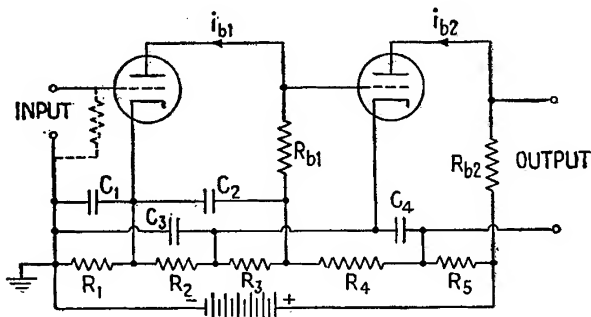


FIG. 5.2.—Loftin-White direct-coupled amplifier.

or greater magnitude than the d-c variations under observation. It is essential, therefore, that the polarizing voltages be closely maintained.

Because of the high cost of maintenance of the battery-operated amplifier, Fig. 5.1, and because of its weight and bulk, it is desirable to utilize a circuit that can be powered from a single battery or from a rectifier. One such circuit is shown in Fig. 5.2. If the input circuit is not continuous, a resistor must be connected across the input terminals to provide a d-c path between grid and cathode.

The polarizing potentials for the tubes are provided by the voltage drops across the resistors R_1, R_2, R_3, R_4, R_5 . The by-pass capacitors, C_1, C_2, C_3, C_4 provide low-impedance paths for alternating currents, thus helping to maintain the polarizing potentials constant.

6. Resistance-coupled Amplifiers.—The resistance-coupled amplifier is probably the most common type of voltage amplifier used for audio and wide-band applications. In a resistance-coupled amplifier the plate of one tube is connected to the grid of the succeeding tube by means of a blocking (or coupling) capacitor. Resistors are placed in the plate and the grid circuits.

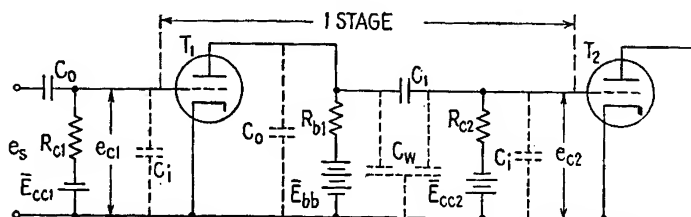


FIG. 6.1.—Resistance-coupled amplifier stage using a triode with batteries for polarizing potentials.

One stage of a resistance-coupled amplifier with its input circuit is shown in Fig. 6.1. The coupling capacitor C_1 prevents or blocks the flow of direct current and hence allows the plate and grid-bias voltages to be independent. The alternating component of the potential developed across the plate-load resistor R_{b1} causes an alternating current to flow in the circuit consisting of R_{c2} and C_1 in series. The a-c potential drop across the grid resistor R_{c2} is applied to the grid of T_2 .

The grid polarizing potentials \bar{E}_{cc1} and \bar{E}_{cc2} are obtained, in many practical applications, from a single battery, and the plate power-supply potential \bar{E}_{bb} is utilized by several stages. Certain precautions are necessary in such cases. These are discussed in Sec. 33.

The triodes in Fig. 6.1 often are replaced by pentodes with which it is possible to obtain higher voltage amplification per stage. It is also desirable in many applications to supply all polarizing potentials from one power source, usually a rectifier and smoothing circuit. A typical circuit diagram of one stage of a resistance-coupled amplifier utilizing pentodes with *cathode* biasing, Sec. 7, is shown in Fig. 6.2. This circuit differs from that of Fig. 6.1 only in the tubes and the means of obtaining the polarizing potentials.

In both Figs. 6.1 and 6.2, C_i represents the input capacitance

to the tube, C_o the output capacitance of the tube, and C_w the distributed capacitance of the interstage wiring. These capacitances are quite small and may be neglected except at relatively high frequencies.

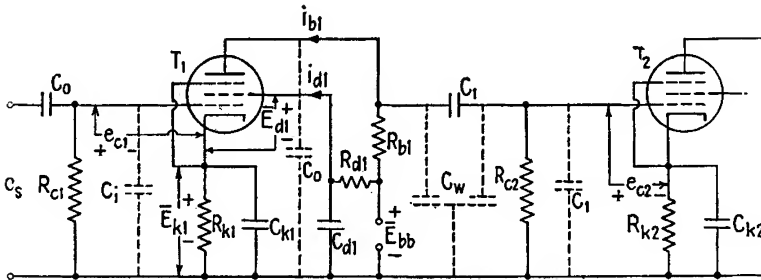


FIG. 6.2.—Resistance-coupled stage using a pentode and cathode biasing.

7. Polarizing Potentials in a Resistance-coupled Stage.—All tube polarizing potentials are referred to the cathode. The grid polarizing potential for the pentode T_1 , Fig. 6.2, is obtained from the voltage drop across the cathode resistor R_{k1} . This method of obtaining the grid-bias voltage is known as *cathode biasing* or *self-biasing*. The voltage drop \bar{E}_{k1} is given by

$$\bar{E}_{k1} = R_{k1}(\bar{I}_{b1} + \bar{I}_{d1}) \quad (7.1)$$

where \bar{I}_{b1} and \bar{I}_{d1} are the d-c components of the plate and screen-grid currents. When the control-grid potential is negative, $\bar{I}_{b1} + \bar{I}_{d1}$ is the total cathode current. If the control grid draws current, this current must be taken into consideration in calculating \bar{E}_{k1} .

When no current flows in R_{c1} , the potential of the control grid with respect to the negative terminal of the power supply \bar{E}_{bb} is zero and the potential of the cathode is $+\bar{E}_{k1}$ with respect to the same reference potential. Therefore, the potential of the control grid with respect to the cathode is $-\bar{E}_{k1}$.

When an alternating potential is applied to the control grid, the plate and screen-grid currents have alternating components. Provision must be made, therefore, for maintaining \bar{E}_k constant if a constant grid bias is desired. The by-pass capacitor C_{k1} is provided for this purpose. At the lowest frequency to be amplified, the reactance of C_{k1} must be negligible when compared with R_{k1} if no variation in the cathode potential is desired.¹ In the

¹ Actually, $Z_{k1}(1 + \mu)$ must be negligible when compared with $(r_p + Z_L)$ for no cathode degeneration; cf. Sec. 31.

absence of C_{k1} or when the reactance of C_{k1} is not negligible, the potential of the cathode changes in the same direction as the signal potential, thus reducing the variation in the potential applied to the control grid. This phenomenon, which is known as *cathode degeneration*, is discussed in Sec. 31.

The screen-grid polarizing potential \bar{E}_{d1} is obtained from the power supply \bar{E}_{bb} through the resistor R_{d1} . \bar{E}_{d1} may be calculated from

$$\bar{E}_{d1} = \bar{E}_{bb} - \bar{E}_{k1} - R_{d1}\bar{I}_{d1} \quad (7.2)$$

It is usually desirable to maintain \bar{E}_{d1} constant. This is accomplished by making the by-pass capacitance C_{d1} large so that its reactance at the lowest frequency to be amplified is negligible compared with the sum of R_{d1} and the internal impedance of the power supply. When the reactance of C_{d1} is not small, *screen-grid degeneration* occurs. For many applications it is sufficient if the reactance of the by-pass capacitor at the lowest frequency for which neither cathode degeneration nor screen-grid degeneration is desired is not greater than one-tenth the resistance by-passed.

The plate polarizing potential may be calculated from

$$\bar{E}_{b1} = \bar{E}_{bb} - \bar{E}_{k1} - \bar{I}_{b1}R_{b1} \quad (7.3)$$

8. Equivalent Plate Circuits of Resistance-coupled Amplifiers.—

The equivalent alternating-current plate circuit for the amplifiers shown in Figs. 6.1 and 6.2 are identical and may be drawn with

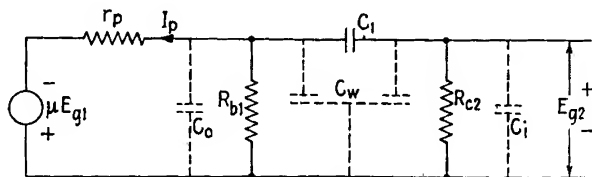


FIG. 8.1.—Equivalent plate circuit using a constant-voltage generator.

either a constant-voltage generator or a constant-current generator. The constant-voltage-generator form of the equivalent plate circuit is shown in Fig. 8.1 and the constant-current-generator form in Fig. 8.2. It should be emphasized that either form of the equivalent circuit can be applied to any amplifier circuit.

A complex expression for the voltage amplification A of the resistance-coupled stage, at any frequency, obtained from either of the equivalent circuits is

$$A = \frac{-\mu}{1 + \frac{r_p}{R_b} + \frac{r_p}{R_c} + \frac{C_c}{C_1} + \frac{r_p C_L}{R_c C_1} + \frac{r_p C_c}{R_b C_1}} \quad (8.1)$$

$$- j \left[\frac{1}{\omega R_c C_1} + \frac{r_p}{\omega R_b R_c C_1} - \frac{\omega r_p C_L C_c}{C_1} - \omega r_p (C_L + C_c) \right]$$

where

$$C_L = C_o + C_w' \quad \text{and} \quad C_c = C_i + C_w'' \quad (8.2)$$

C_w' being that part of the distributed capacitance of the interstage wiring which is in shunt with R_{b1} and C_w'' the similar capacitance in shunt with R_{c2} .

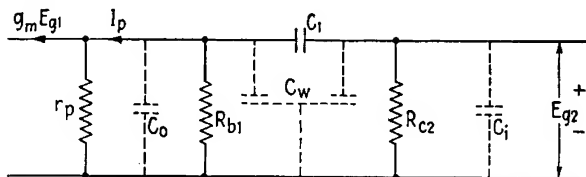


FIG. 8.2.—Equivalent plate circuit using a constant-current generator.

Equation (8.1) is so complicated that the problem is usually divided into three frequency ranges, depending upon the magnitude of the reactances of C_1 and C_s , where

$$C_s = C_o + C_w + C_i \quad (8.3)$$

The *low-frequency* range is that range of frequencies for which the reactance of the coupling capacitor C_1 is not negligible when compared with R_{c2} and the a-c voltage developed across R_{c2} is appreciably less than that developed across the plate-load impedance.

In the *mid-frequency* range (if it exists) the reactance of C_1 is negligibly small compared with R_{c2} , and the reactance of C_s is very large compared with the parallel combination of R_{b1} and R_{c2} .

The *high-frequency* range is defined as that range of frequencies for which the reactance of C_s is small enough to cause appreciable shunting effect. As in the mid-frequency range, the reactance of C_1 is negligible at high frequencies.

The expression for the voltage amplification in each of these frequency ranges can be obtained from (8.1), but the development of the expressions from the simplified equivalent circuits is instructive.

9. Voltage Amplification at Mid-frequencies.—The equivalent plate circuit at mid-frequencies is shown in Fig. 9.1, in which the

numerical subscripts have been omitted from the plate-load and grid resistors. The grid resistor R_c is to be associated with the grid circuit of the following stage.

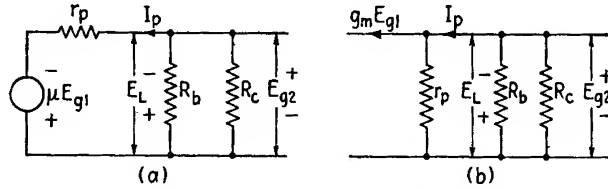


FIG. 9.1.—Equivalent plate circuits of a resistance-coupled stage for the mid-frequency range.

As shown in Fig. 9.1a, the plate-load impedance at mid-frequencies is resistive. The voltage amplification is

$$A_m = \frac{E_{g2}}{E_{g1}} = \frac{-\mu R_L}{r_p + R_L} \quad (9.1)$$

where the subscript m denotes the mid-frequency range and

$$R_L = \frac{R_b R_c}{R_b + R_c} \quad (9.2)$$

Equation (9.1) may be rewritten as

$$A_m = \frac{-\mu}{1 + \frac{r_p}{R_b} + \frac{r_p}{R_c}} \quad (9.3)$$

by dividing both numerator and denominator by R_L and substituting for R_L from (9.2), or as

$$A_m = \frac{-g_m R_b}{1 + \frac{R_b}{r_p} + \frac{R_b}{R_c}} \quad (9.4)$$

by multiplying both numerator and denominator of (9.3) by R_b/r_p .

Similarly, if R_L' is defined by

$$\frac{1}{R_L'} = \frac{1}{r_p} + \frac{1}{R_b} + \frac{1}{R_c} \quad (9.5)$$

the mid-frequency voltage amplification, from either (9.4) or directly from Fig. 9.1b, is

$$A_m = -g_m R_L' \quad (9.6)$$

or A_m is given by the product of the transconductance and the equivalent of r_p , R_b , R_c , in parallel.

It follows that

$$A_m = \frac{\mu}{1 + \frac{r_p}{R_b} + \frac{r_p}{R_c}} \angle 180^\circ = \frac{g_m R_b}{1 + \frac{R_b}{r_p} + \frac{R_b}{R_c}} \angle 180^\circ = g_m R_L' \angle 180^\circ \quad (9.7)$$

These relations indicate that the magnitude of the voltage amplification in the mid-frequency range is independent of frequency and that the relative phase angle is zero.

10. Voltage Amplification at Low Frequencies.—Since the presence of C_o , C_w , C_i may be neglected at low frequencies, the equivalent plate circuits with constant-voltage and constant-current generators are as shown in Figs. 10.1a and b.

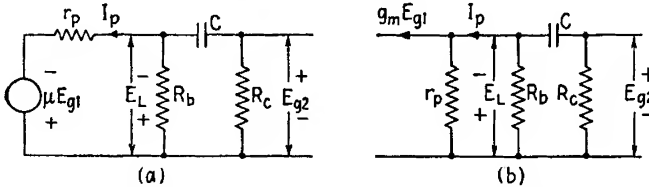


FIG. 10.1.—Equivalent plate circuits of a resistance-coupled stage for the low-frequency range.

The coupling capacitor C and the grid resistor R_c form a potential divider across E_L , and hence the alternating potential E_{g2} applied to the grid of the succeeding tube is

$$E_{g2} = \frac{-R_c}{Z_2} E_L \quad (10.1)$$

where

$$Z_2 = R_c - \frac{j}{\omega C} \quad (10.2)$$

and the voltage amplification is

$$A_l = \frac{E_L}{E_{g1}} \frac{E_{g2}}{E_L} \quad (10.3)$$

the subscript l denoting the low-frequency range. From the equivalent circuit of Fig. 10.1a,

$$\frac{E_L}{E_{g1}} = \frac{\mu R_b Z_2}{r_p(R_b + Z_2) + R_b Z_2} \quad (10.4)$$

Substituting in (10.3) from (10.1) and (10.4),

$$A_l = \frac{-\mu R_b R_c}{r_p(R_b + Z_2) + R_b Z_2} \quad (10.5)$$

By substituting for Z_2 in (10.5), simplifying the denominator, and dividing both numerator and denominator by $R_b R_c$, the complex expression for the voltage amplification becomes

$$A_i = \frac{-\mu}{1 + \frac{r_p}{R_b} + \frac{r_p}{R_c} - j \frac{1}{\omega C R_c} \left(1 + \frac{r_p}{R_b}\right)} \quad (10.6)$$

Equation (10.6) can be converted to

$$A_i = \frac{-g_m R_b}{1 + \frac{R_b}{r_p} + \frac{R_b}{R_c} - j \frac{1}{\omega C R_c} \left(1 + \frac{R_b}{r_p}\right)} \quad (10.7)$$

by multiplying both numerator and denominator by R_b/r_p .

An alternate expression, often useful, may be obtained for the low-frequency voltage amplification either by dividing both the numerator and denominator of (10.7) by R_b or by derivation directly from Fig. 10.1b. This equation is

$$A_i = \frac{-g_m R_L'}{1 - j \frac{R_L'}{\omega C R_c R'}} \quad (10.8)$$

where

$$R' = \frac{r_p R_b}{r_p + R_b} \quad (10.9)$$

is the equivalent of r_p and R_b , in parallel, and R_L' defined by (9.5), is the equivalent of r_p , R_b , R_c , in parallel. However, the total resistance of the circuit of Fig. 10.1b as measured from the terminals of C is

$$R_i = R_c + R' = \frac{R_c R'}{R_L'} \quad (10.10)$$

and (10.8) becomes

$$A_i = \frac{-g_m R_L'}{1 - j \frac{1}{\omega C R_i}} \quad (10.11)$$

The magnitude of the voltage amplification at low frequencies, therefore, is either

$$|A_i| = \frac{\mu}{\sqrt{\left(1 + \frac{r_p}{R_b} + \frac{r_p}{R_c}\right)^2 + \frac{1}{\omega^2 C^2 R_c^2} \left(1 + \frac{r_p}{R_b}\right)^2}} \quad (10.12)$$

from (10.6), or

$$|A_i| = \frac{g_m R_b}{\sqrt{\left(1 + \frac{R_b}{r_p} + \frac{R_b}{R_c}\right)^2 + \frac{1}{\omega^2 C^2 R_c^2} \left(1 + \frac{R_b}{r_p}\right)^2}} \quad (10.13)$$

from (10.7); and from (10.11)

$$|A_i| = \frac{g_m R_L'}{\sqrt{1 + \frac{1}{\omega^2 C^2 R_i^2}}} \quad (10.14)$$

The relative phase angle θ_i is expressed by

$$\tan \theta_i = + \frac{1 + \frac{r_p}{R_b}}{\left(1 + \frac{r_p}{R_b} + \frac{r_p}{R_c}\right) R_c \omega C} \quad (10.15)$$

or

$$\tan \theta_i = \frac{1 + \frac{R_b}{r_p}}{\left(1 + \frac{R_b}{r_p} + \frac{R_b}{R_c}\right) R_c \omega C} \quad (10.16)$$

or

$$\tan \theta_i = \frac{1}{\omega C R_i} \quad (10.17)$$

The expressions (10.12), (10.13), (10.14) show that the variation in $|A|$ as a function of frequency in the low-frequency range is caused by the presence of the coupling capacitor C and that high voltage amplification is obtained by making R_b and R_c large as compared with r_p and also by making the time constant CR_c large as compared with the period of the voltage to be amplified.

The variational characteristics μ , g_m , and r_p of the tube must be determined for the polarizing potentials at which the tube is operated.

The frequency in the low-frequency band at which the power delivered to the resistor R_c is one-half that delivered at frequencies in the mid-frequency band, the input voltage E_{g1} being of constant amplitude, is called the *lower half-power* frequency. This frequency may also be thought of as that frequency at which either the voltage amplification reduces to 0.707 of the mid-frequency value or the gain is 3 db below the mid-frequency gain. The lower half-power frequency is designated by f' .

$|A_i|$ reduces to 0.707 of its mid-frequency value when

$$\omega' = \frac{1 + \frac{r_p}{R_b}}{CR_c \left(1 + \frac{r_p}{R_b} + \frac{r_p}{R_c}\right)} \quad (10.18)$$

or

$$\omega' = \frac{1 + \frac{R_b}{r_p}}{CR_c \left(1 + \frac{R_b}{r_p} + \frac{R_b}{R_c}\right)} \quad (10.19)$$

or

$$\frac{1}{\omega' C} = R_i \quad (10.20)$$

Substituting from (10.18), (10.19), (10.20), in (10.6), (10.7), (10.11), and in (10.15), (10.16), (10.17),

$$A_i = \frac{A_m}{1 - j \frac{f'}{f}} \quad (10.21)$$

from which

$$|A_i| = \frac{|A_m|}{\sqrt{1 + (f'/f)^2}} \quad (10.22)$$

and

$$\tan \theta_i = + \frac{f'}{f} \quad (10.23)$$

where f is the frequency of the applied signal.

The relative phase angle θ_i as given by (10.15), (10.16), (10.17), and (10.23) is $+45^\circ$ at the lower half-power frequency.

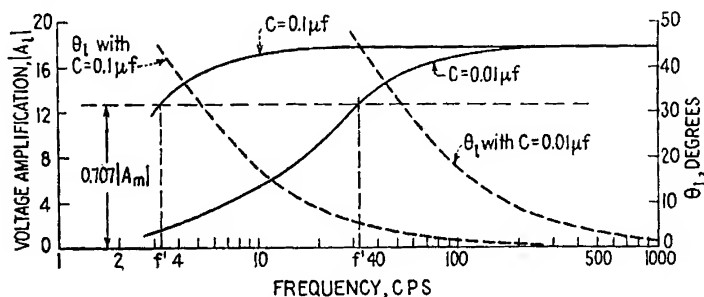


FIG. 10.2.—Low-frequency response and relative phase angle of a resistance-coupled stage.

The low-frequency response and relative phase angle of a resistance-coupled stage employing a triode for which $\mu = 20$, $g_m = 2,000 \mu\text{mhos}$, $r_p = 10,000 \text{ ohms}$, and with $R_b = 100,000 \text{ ohms}$

and $R_c = 0.5$ megohm is shown in Fig. 10.2 for two values of C of $0.1 \mu\text{f}$ and $0.01 \mu\text{f}$.

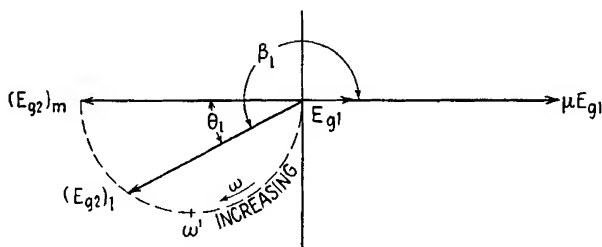


FIG. 10.3.—Polar relation of E_{g2} to E_{g1} at low frequencies in a resistance-coupled stage.

The mid-frequency voltage amplification of this stage is 17.86, and the lower half-power frequencies for each of the values of C are shown. It will be noted from (10.18), (10.19), (10.20) that the value of R_b has very little effect on the value of f' but that the product of $R_c C$ is of major importance. These results are shown in the polar diagram of Fig. 10.3, where the relative phase angle θ_1 and the phase β_1 of E_{g2} with respect to E_{g1} are shown.

11. Voltage Amplification at High Frequencies.—The high-frequency range of a resistance-coupled amplifier is that range in which C_s has appreciable shunting effect and the reactance of C is negligible as compared with R_c . The equivalent plate circuits at high frequencies are shown in Fig. 11.1.

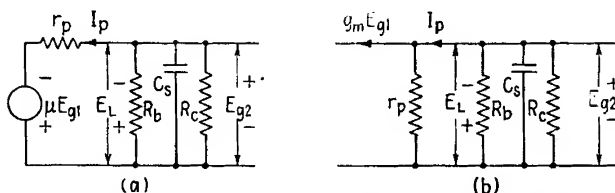


FIG. 11.1.—Equivalent plate circuits of a resistance-coupled stage for the high-frequency range.

In the high-frequency range the voltage developed across R_c is the total voltage developed across the plate-load impedance Z_L , and

$$A_k = \frac{-\mu Z_L}{r_p + Z_L} = \frac{-g_m Z_L}{1 + \frac{Z_L}{r_p}} \quad (11.1)$$

From Fig. 11.1a, Z_L is given by

$$Z_L = \frac{R_L}{1 + j\omega C_s R_L} \quad (11.2)$$

where

$$R_L = \frac{R_b R_c}{R_b + R_c} \quad (11.3)$$

and

$$C_s = C_o + C_w + C_i \quad (11.4)$$

It follows that

$$A_h = \frac{-\mu}{1 + \frac{r_p}{R_b} + \frac{r_p}{R_c} + j\omega C_s r_p} \quad (11.5)$$

or

$$A_h = \frac{-g_m R_b}{1 + \frac{R_b}{r_p} + \frac{R_b}{R_c} + j\omega C_s R_b} \quad (11.6)$$

and

$$A_h = \frac{-g_m R_L'}{1 + j\omega C_s R_L'} \quad (11.7)$$

where R_L' is the equivalent of r_p , R_b , R_c , in parallel, (9.5).

The upper half-power frequency f'' is given by

$$\omega'' = \frac{1 + \frac{r_p}{R_b} + \frac{r_p}{R_c}}{C_s r_p} \quad (11.8)$$

$$\omega'' = \frac{1 + \frac{R_b}{r_p} + \frac{R_b}{R_c}}{C_s R_b} \quad (11.9)$$

or

$$\frac{1}{\omega'' C_s} = R_L' \quad (11.10)$$

As R_b decreases, ω'' increases and $|A_m|$ decreases.

By combining the corresponding equations of the groups (11.5), (11.6), (11.7) and (11.8), (11.9), (11.10),

$$A_h = \frac{A_m}{1 + j \frac{f}{f''}} \quad (11.11)$$

The magnitude of A_h is

$$|A_h| = \frac{\mu}{\sqrt{\left(1 + \frac{r_p}{R_b} + \frac{r_p}{R_c}\right)^2 + \omega^2 C_s^2 r_p^2}} \quad (11.12)$$

$$|A_h| = \frac{g_m R_b}{\sqrt{\left(1 + \frac{R_b}{r_p} + \frac{R_b}{R_c}\right)^2 + \omega^2 C_s^2 R_b^2}} \quad (11.13)$$

and

$$|A_h| = \frac{|A_m|}{\sqrt{1 + (f/f'')^2}} \quad (11.14)$$

The relative phase angle θ_h is given by

$$\tan \theta_h = - \frac{\omega C_s r_p}{1 + \frac{r_p}{R_b} + \frac{r_p}{R_c}} \quad (11.15)$$

$$\tan \theta_h = - \frac{\omega C_s R_b}{1 + \frac{R_b}{R_c} + \frac{R_b}{r_p}} \quad (11.16)$$

or

$$\tan \theta_h = - \frac{f}{f''} \quad (11.17)$$

Thus, at the upper half-power frequency θ_h is -45° .

These considerations show that, when the terms involving C_s are negligible, the expressions become identical with those for the mid-frequencies.

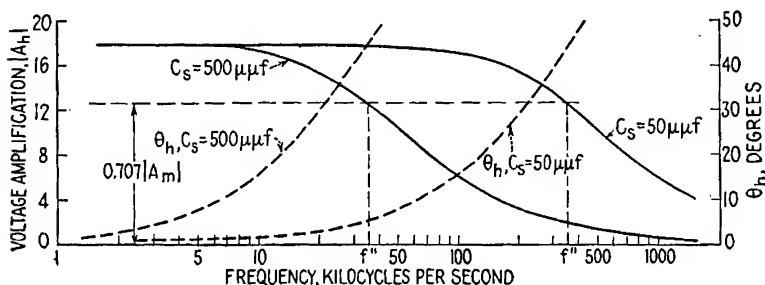


FIG. 11.2.—High-frequency response of a resistance-coupled amplifier stage.

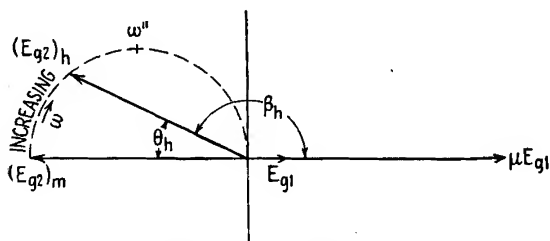


FIG. 11.3.—Polar relation of E_{q2} to E_{q1} at high frequencies in a resistance-coupled stage.

The high-frequency response and relative phase angle θ_h of the stage used in Fig. 10.2 are shown in Fig. 11.2 for values of C_s of 50 and 500 μf .

The upper half-power frequencies for each value of C_s are found from (11.8), (11.9), or (11.10) to be 356,500 cps when C_s is 50 $\mu\mu\text{f}$ and 35,650 cps when C_s is 500 $\mu\mu\text{f}$. These half-power frequencies are shown in Fig. 11.2.

The polar diagram for high-frequency performance of a resistance-coupled stage is given in Fig. 11.3.

12. Special Case, $R_b \ll r_p$ and $R_b \ll R_c$.—In using pentodes r_p is of the order of 1 megohm and R_c is usually above 250,000 ohms. The plate-load resistor R_b is often as low as a few thousand ohms when wide-band operation is desired. When $R_b \ll r_p$ and $R_b \ll R_c$,

$$R_L' \doteq R_L \doteq R_b \quad (12.1)$$

and, therefore, from either (9.4) or (9.6),

$$A_m \doteq -g_m R_b \quad (12.2)$$

from (10.18) to (10.20),

$$\omega' \doteq \frac{1}{CR_c} \quad (12.3)$$

from (11.8) to (11.10),

$$\omega'' \doteq \frac{1}{C_s R_b} \quad (12.4)$$

and (10.22) and (11.11) reduce to

$$A_t \doteq \frac{-g_m R_b}{1 - j \frac{f'}{f}} \quad (12.5)$$

and

$$A_h \doteq \frac{-g_m R_b}{1 + j \frac{f}{f''}} \quad (12.6)$$

13. Universal Amplification Curve for a Resistance-coupled Stage.—From (10.22) the relative voltage amplification at low frequencies is

$$\frac{|A_t|}{|A_m|} = \frac{1}{\sqrt{1 + (f'/f)^2}} \quad (13.1)$$

and at high frequencies, from (11.14), the relative voltage amplification is

$$\frac{|A_h|}{|A_m|} = \frac{1}{\sqrt{1 + (f/f'')^2}} \quad (13.2)$$

The relative phase angles for low and high frequencies are given

by (10.23) and (11.17). Because of the similarity of (13.1) and (13.2) and of (10.23) and (11.17), it is convenient to indicate the relative amplification and relative phase angle at low as well as at high frequencies by the curves shown in Fig. 13.1.

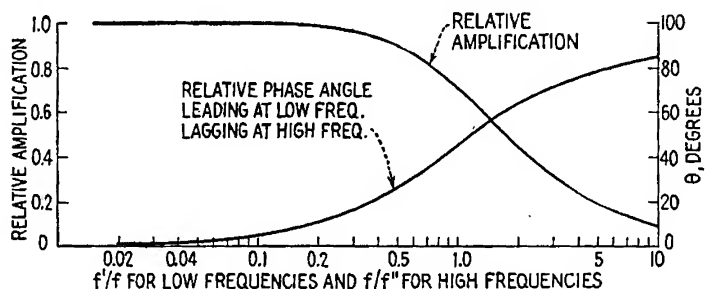


FIG. 13.1.—Universal amplification and phase curves for resistance-coupled amplifier stages.

Example.—The results shown in Figs. 10.2 and 11.2 may be obtained by use of the universal amplification curve. In this example $C = 0.01 \mu\text{f}$ and $C_s = 500 \mu\text{f}$.

For the low-frequency case,

$$f' = \frac{1 + \frac{r_p}{R_b}}{2\pi R_c C \left(1 + \frac{r_p}{R_b} + \frac{r_p}{R_c}\right)} = \frac{1.1}{0.01\pi \cdot 1.12} = 31.2 \text{ cps}$$

In order to determine the relative amplification and relative phase angle at 60 cps, the ratio f'/f is determined as

$$\frac{f'}{f} = \frac{31.2}{60} \doteq 0.52$$

It is found from the curves that, when $f'/f = 0.52$, $\theta \doteq 28^\circ$ and the relative amplification is 0.89. Since it is shown in Sec. 10 that the mid-frequency voltage amplification of this stage is 17.86, the voltage amplification at 60 cps is

$$|A| = 0.89 \cdot 17.86 = 15.9$$

At high frequencies

$$f'' = \frac{\left(1 + \frac{r_p}{R_b} + \frac{r_p}{R_c}\right)}{2\pi r_p C_s} = \frac{1.12}{\pi \cdot 10^{-6}} = 35,650 \text{ cps}$$

The frequency ratio when $f = 20,000$ cps is

$$\frac{f}{f''} = 0.561$$

One obtains from the high-frequency curve of Fig. 13.1 a relative amplification of 0.87 and a relative phase angle of approximately -30° . The voltage amplification at 20 keps is $0.87 \cdot 17.86 = 15.5$, and the relative phase angle is -30° .

The time constant CR_c is very important for transient response. This is discussed in detail in Chap. VI.

15. Effects of Variations in R_b and \bar{E}_{bb} .—The effects of variations in the coupling capacitance on the low-frequency response and of variations in the shunting capacitance on the high-frequency response have been discussed in Secs. 10 and 11. The variations in the over-all frequency response caused by changing independently the plate-load resistance and the potential of the plate power supply are important.

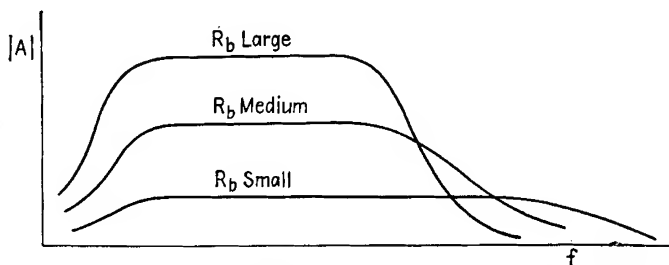


FIG. 15.1.—Effect on the voltage amplification of varying R_b . \bar{E}_{bb} is constant.

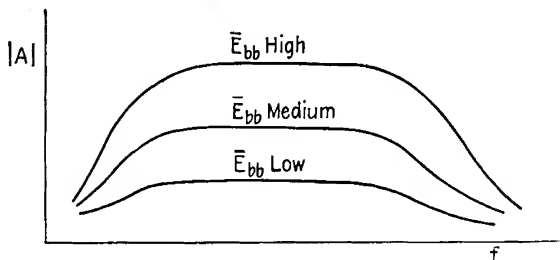


FIG. 15.2.—Effect on the voltage amplification of varying \bar{E}_{bb} . R_b is constant.

When the plate-load resistance is high, the quiescent plate current of the tube is low and its transconductance is small. A decrease in R_b results in larger plate current and increased transconductance. Since the increase in transconductance is always proportionally smaller than the decrease in R_b , the resulting voltage amplification in the low-frequency and mid-frequency ranges is smaller the lower the value of R_b . The voltage amplification in the upper part of the high-frequency range may be increased, however, by reducing the value of R_b . With low values of R_b , the mid-frequency range is extended to higher frequencies. With high values of R_b , the voltage amplification is increased in the low-frequency and mid-frequency ranges and the upper half-power frequency is lowered, Fig. 15.1.

The curves in Fig. 15.2 show the effect of varying the potential \bar{E}_{bb} of the plate power supply, R_b being kept constant. The voltage amplification is increased appreciably by using higher values of \bar{E}_{bb} , as a result of the increase in plate current and the associated increase in transconductance. There is no appreciable change in the lower and upper half-power frequencies when \bar{E}_{bb} is changed, the circuit constants remaining unchanged.

16. Impedance-coupled Amplifiers.—Impedance-coupled amplifiers are amplifiers in which either R_b or R_c , or both, of the resistance-coupled amplifier, are replaced by an inductor. The inductor is usually an iron-core coil, called an *audio-frequency choke*, which has

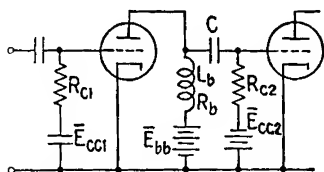


FIG. 16.1.—One impedance-coupled amplifier stage.

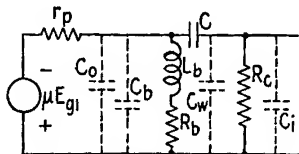


FIG. 16.2.—Equivalent plate circuit of one impedance-coupled amplifier stage.

large inductance and relatively small resistance. This type of coupling is seldom used except in limited applications where large voltage amplification over a limited frequency range is desired. The circuit diagram of one stage of an impedance-coupled amplifier and its equivalent plate circuit are shown in Figs. 16.1 and 16.2.

The frequency response of an impedance-coupled amplifier is poorer at both low and high frequencies than for a resistance-coupled amplifier. The decrease in voltage amplification at low frequencies is caused by the decrease in the impedance of the inductor in this frequency range. The decrease in voltage amplification at high frequencies results from the appreciable increase in C_s due to the distributed capacitance of the inductor (C_b , Fig. 16.2). Each of these defects reduces the mid-frequency range.

The equations for the voltage amplification of an impedance-coupled amplifier can be obtained from an analysis of the circuit of Fig. 16.2, the same procedure being followed as with the resistance-coupled amplifier.

17. Wide-band (Video) Amplifiers.—The amplification of voltages for audio purposes is simpler than for some other applications since the audio-frequency range does not exceed 20 kcps and the ear does not distinguish small relative phase shifts of the component frequencies of a composite sound. On the other hand,

the amplification of voltages that are to be applied to the deflecting plates of a cathode-ray tube must be performed in such manner that the wave shape of the signal voltage is unaltered. It is shown in Chaps. IX and XVIII that for the exact reproduction of any waveform the circuit must introduce neither frequency nor phase distortion. Since this ideal is impossible to attain in amplifiers, approximations are used that result in wave-shape reproductions satisfactory for most applications. It is assumed that operation of the tubes of the amplifier on the linear portion of their characteristics satisfies the condition of no nonlinear distortion.

The video signal utilized in television, or a saw-tooth voltage, or a rectangular voltage, or any voltage in which there are discontinuities requires wide-band response as well as no phase distortion. The requirements of frequency response placed on any amplifier are determined by the type of signal to be reproduced and the accuracy with which it must be reproduced. In the video amplifier of television apparatus, for example, the frequency range should extend from as low as 20 cps to approximately 5 mcps. The broader the response band of the equipment, the sharper the television picture. It is quite possible, however, to obtain satisfactory results for other types of cathode-ray-tube applications with upper half-power frequencies of much less than 5 mcps. For example, one type of general-purpose cathode-ray oscilloscope is designed so that its frequency response is constant to within ± 1 db up to 100,000 cps.

The relative phase shift should be proportional to the frequency of the signal component for all frequencies within the pass band. Another way of stating this requirement is that the amplifier should have constant time delay over the pass band. It has been determined empirically that an amplifier must have no frequency distortion within the frequency range extending from the frequency equal to one-tenth the lowest rate of recurrence to the frequency equal to ten times the highest rate of recurrence of any rectangular wave to be reproduced.

It is shown in Sec. 11 that the upper half-power frequency of a resistance-coupled amplifier can be made much larger than required for most applications by reducing the value of the plate-load resistance R_b and by reducing the shunting capacitance. The shunting capacitance may be made small by choosing tubes having small input and output capacitances, by using circuit elements with small physical dimensions, which are placed so as to have low

capacitance to the chassis, and by wiring the amplifier with great care. However, the voltage amplification is low when small values of R_b are used. Also, at high frequencies the *tangent* of the relative phase angle, not the relative phase angle itself, is proportional to the frequency in a resistance-coupled amplifier, Sec. 11. Therefore, a resistance-coupled amplifier does not always meet the requirements for a broad-band amplifier.

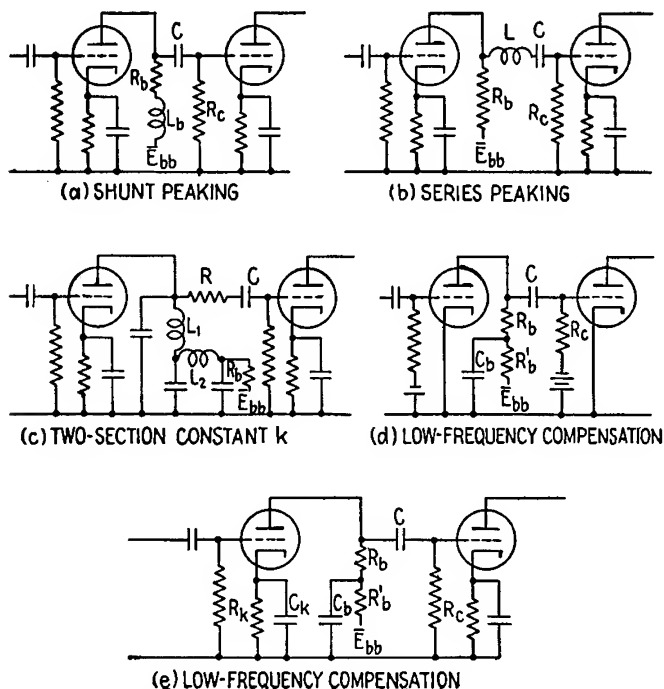


FIG. 17.1.—Typical high-frequency (a, b, c) and low-frequency (d, e) compensating circuits.

The frequency response and phase shift in a resistance-coupled amplifier can be improved by the use of compensating circuits. The circuits a, b, c, Fig. 17.1, are three of the large number of possible high-frequency compensating circuits. The discussion in this presentation is limited to shunt-peaking compensation, which is probably the most commonly used, though more desirable results for certain applications can be obtained by other types. Low-frequency compensation is discussed briefly.

18. Shunt-peaking Compensation at High Frequencies.—The circuit diagram of one stage of a shunt-peaking compensated

resistance-coupled amplifier is shown in Fig. 18.1a. The output capacitance C_o of the first tube, the interstage wiring capacitance C_w , the input capacitance C_i of the second tube are indicated by the broken lines. L_b is the compensating inductance. The equivalent plate circuit at high frequencies is shown in Fig. 18.1b.

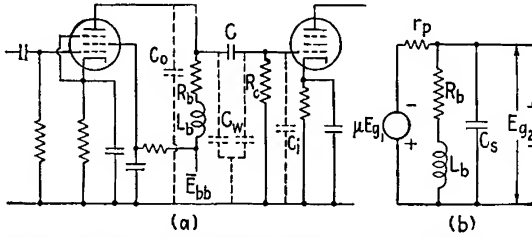


FIG. 18.1.—A single-stage shunt-peaking amplifier and its high-frequency equivalent plate circuit.

It is assumed that $R_c \gg R_b$ and that $r_p \gg R_b$. Z_L is the equivalent impedance of the two parallel branches, C_s , and R_b in series with L_b . Since $Z_L \ll r_p$, the complex expression for the voltage amplification at high frequencies is

$$A \doteq -g_m Z_L \quad (18.1)$$

The expression for Z_L is

$$Z_L = \frac{(R_b + j\omega L_b)(-j/\omega C_s)}{R_b + j\left(\omega L_b - \frac{1}{\omega C_s}\right)} \quad (18.2)$$

which, by multiplying both numerator and denominator by $j\omega C_s$, becomes

$$Z_L = \frac{R_b + j\omega L_b}{(1 - \omega^2 C_s L_b) + j\omega C_s R_b} \quad (18.3)$$

Rationalizing (18.3),

$$Z_L = \frac{R_b - j(\omega C_s R_b^2 - \omega L_b + \omega^3 C_s L_b^2)}{(1 - \omega^2 C_s L_b)^2 + \omega^2 C_s^2 R_b^2} \quad (18.4)$$

Let N be the ratio of the reactance of the compensating inductance L_b at the upper half-power frequency of the uncompensated stage to the plate-load resistance R_b ,

$$N = \frac{\omega'' L_b}{R_b} \quad (18.5)$$

The expression for Z_L in terms of R_b , N , f , and f'' may be obtained

by substituting in (18.4) the expression for L_b from (18.5) and C_s from (12.4),

$$Z_L = R_b \left\{ \frac{1 - j \left[(1 - N) \frac{f}{f''} + N^2 \left(\frac{f}{f''} \right)^3 \right]}{[1 - N(f/f'')^2]^2 + (f/f'')^2} \right\} \quad (18.6)$$

The expression for the voltage amplification of the compensated stage is, therefore,

$$A \doteq -g_m R_b \left\{ \frac{1 - j \left[(1 - N) \frac{f}{f''} + N^2 \left(\frac{f}{f''} \right)^3 \right]}{[1 - N(f/f'')^2]^2 + (f/f'')^2} \right\} \quad (18.7)$$

The magnitude of the voltage amplification is

$$|A| \doteq g_m R_b \frac{\sqrt{1 + \left[(1 - N) \frac{f}{f''} + N^2 \left(\frac{f}{f''} \right)^3 \right]^2}}{[1 - N(f/f'')^2]^2 + (f/f'')^2} \quad (18.8)$$

and the relative phase angle is given by

$$\tan \theta \doteq - \left[(1 - N) \frac{f}{f''} + N^2 \left(\frac{f}{f''} \right)^3 \right] \quad (18.9)$$

The value of N for which $|A|$ at the uncompensated upper half-power frequency is equal to its mid-frequency value $g_m R_b$

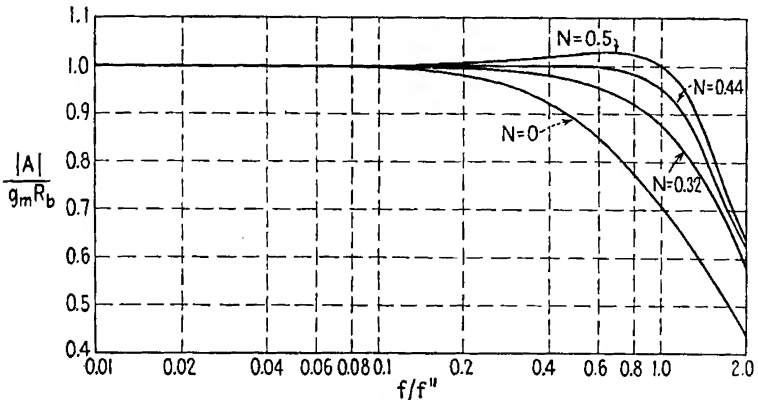


FIG. 18.2.—High-frequency response curves for shunt-peaking compensated resistance-coupled amplifier stages.

may be obtained by setting $f/f'' = 1$ and $|A| = g_m R_b$ in (18.8) and is $N = 0.5$. However, this value does not ensure uniform voltage amplification at frequencies below f'' . The results plotted

in Fig. 18.2 illustrate the variation in $|A|$ with frequency. The compensation obtained when $N = 0.5$ is generally referred to as the compensation for "best frequency response," since the voltage amplification does not deviate appreciably from the mid-frequency value for applied high frequencies up to and including the upper half-power frequency for the uncompensated stage.

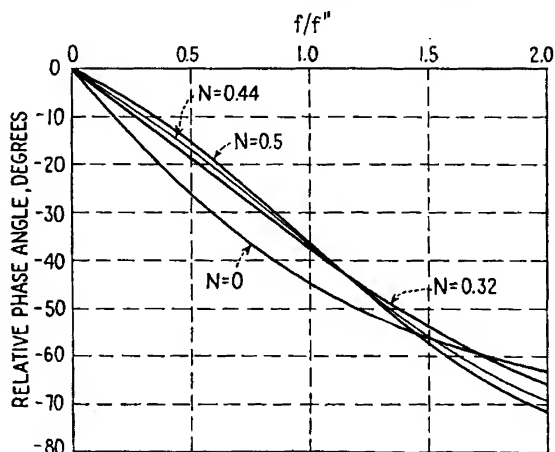


FIG. 18.3.—Relative phase angles at high frequencies in shunt-peaking compensate resistance-coupled amplifier stages.

Figure 18.3 shows with $N = 0.5$ that θ_h is not proportional to frequency within the pass band. This may be illustrated by comparing the relative phase angle for f/f'' equal to 0.1 and 1.0. The value of θ_h when f/f'' is 0.1 is $-2^\circ 53'$. The desired value of θ_h when $f/f'' = 1$ is given by multiplying $-2^\circ 53'$ by 10. This product is $-28^\circ 50'$ rather than $-36^\circ 53'$ as indicated in Fig. 18.3 for the relative phase angle at $f/f'' = 1$. The value of N for constant time delay can be determined from (18.9) and the relation

$$10 \tan^{-1} [(1 - N) \cdot 0.1 + 0.001N^2] = \tan^{-1} (1 - N + N^2)$$

the left-hand member being ten times the relative phase angle when $f/f'' = 0.1$ and the right-hand member the relative phase angle when $f/f'' = 1$. The solution of this equation gives $N = 0.32$, which should be used when no phase distortion is desired.

The relative phase angles are plotted as function of f/f'' in Fig. 18.3 for $N = 0, 0.32, 0.44, 0.5$. The voltage amplification at the uncompensated upper half-power frequency is 87 per cent of its mid-frequency value when $N = 0.32$. Since this decrease in $|A|$ is too large for most applications, a compromise value

$N = 0.44$ should be used when it is desired that both frequency response and time delay be improved simultaneously by means of shunt-peaking compensation. The relative voltage amplification and the relative phase angle are plotted, Figs. 18.2 and 18.3, as functions of f/f'' for $N = 0.44$. With this value of N , the voltage amplification at the uncompensated upper half-power frequency is 95.3 per cent of its mid-frequency value.

19. Figure of Merit.—The mid-frequency voltage amplification of a shunt-peaking compensated amplifier is

$$|A_m| \doteq g_m R_b \quad (19.1)$$

or

$$|A_m| \doteq \frac{g_m}{C_s \omega''} \quad (19.2)$$

when expressed in terms of the uncompensated upper half-power frequency. It is desirable, however, for both ω'' and $|A_m|$ to be large. If $|A_m|$ is multiplied by ω'' , a performance factor

$$F_a = \frac{g_m}{C_s} = \frac{g_m}{C_o + C_w + C_i} \quad (19.3)$$

is obtained for the stage. This performance factor is called the

TABLE 19.1.—FIGURES OF MERIT OF TYPICAL TUBES

Tube	g_m , μmhos	C_o , $\mu\mu\text{f}$	C_i , $\mu\mu\text{f}$	$F_t \cdot 10^6$
6AB7	5,000	5	8	385
6AC7	9,000	5	11	562
6AG7	7,700	7.5	12.5	385
6C5	2,000	25	11	56*
6K7	1,650	12	7	87
6L6	5,200	12	10	236
6S7-G	1,750	8.0	4.4	141
6SF5	1,500	6.4	3.6	50*
6SJ7	1,650	7	6	127
6SK7	2,000	7	6	154
6U7-G	1,600	9	5	114
6V6	4,100	11	10	195
6W7-G	1,225	8.5	5	91
6Y6-G	7,100	8	15	309
6SG7	4,700	7	8.5	303
6SS7	1,930	7	5.5	154
7G7	4,500	7	9	281
7H7	3,800	7	8	253

* The input capacitance of triodes depends upon the plate-circuit impedance. These values assume a voltage amplification of 10 with pure resistive loads.

figure of merit of the stage. C_w is determined by the interstage wiring capacitance, while g_m , C_o , and C_i are determined by the tubes used in the stage. It is convenient, therefore, to compare tubes in terms of a figure of merit that is defined by

$$F_i = \frac{g_m}{C_o + C_i} \quad (19.4)$$

where C_o and C_i represent the output and input capacitances of the tube in question. Even though a given type of tube may not be used in every stage of the amplifier, its figure of merit is of importance. The figures of merit of several tubes are given in Table 19.1. In each case the magnitude of F_i has been multiplied by 10^6 .

20. Example of High-frequency Design Procedure of a Single Stage.—In this example it is assumed that all signals are of such magnitude that nonlinear distortion does not exist. It is necessary that the specifications for frequency response, phase relations, and voltage amplification be available before the design can be accomplished. If the desired frequency response and voltage amplification cannot be obtained with a single stage, more stages must be utilized.

It is assumed further that the required upper half-power frequency f'' for each stage is 3 mcps and that a voltage amplification greater than 12 per stage is sufficient.

The 6AC7 tube is chosen because of its large figure of merit. This pentode has variational characteristics of $r_p = 0.75$ megohm and $g_m = 9,000$ μ mhos when operated with normal polarizing potentials. With this tube a value of $R_c = 0.5$ megohm is satisfactory, and $C = 0.01$ μ f is chosen because of low-frequency considerations.

The amplifier is constructed in a manner such that wiring capacitances are minimized. The coupling capacitor C and also R_b and R_c should be located away from the chassis, but not so far away that appreciable feedback is present. The signal leads should be short, and the by-pass capacitors should be mounted as close to the tube socket as possible. It is necessary to place small paper or mica capacitors in parallel with the electrolytic by-pass capacitors in order effectively to by-pass the high frequencies. These small capacitors should be placed near the tube socket, but the large electrolytic capacitor may be placed in a more convenient location on the chassis.

Since it is not feasible to calculate the capacitance of the inter-stage wiring, it must be determined experimentally. This is accomplished readily by determining the upper half-power frequency of the uncompensated stage either by using a vacuum-tube voltmeter whose input capacitance is known or by using the tube of the following stage as a detector. Any value of R_b may be used for this purpose. It is desirable, however, that a value be used which is of the same order of magnitude as that required in the amplifier. The test value of R_b used in this example is $R_{\text{test}} = 5,000$ ohms. It is determined, for example, that the upper half-power frequency f_{test}'' with the 5,000-ohm plate-load resistor is 1 mcps.

Since $R_{\text{test}} \ll r_p$ and $R_{\text{test}} \ll R_c$, C_s is calculated from (12.4)

$$C_s = \frac{1}{2\pi \cdot 10^6 \cdot 5,000} = 32 \cdot 10^{-12} \text{ farad} = 32 \mu\text{f}$$

The value of R_b for $f'' = 3 \cdot 10^6$ cps, the required upper half-power frequency of the uncompensated stage, is given by

$$R_b = \frac{1}{2\pi \cdot 3 \cdot 10^6 \cdot 32 \cdot 10^{-12}} = 1,667 \text{ ohms}$$

Combining (12.4) with the expression

$$C_s = \frac{1}{2\pi f_{\text{test}}'' R_{\text{test}}} \quad (20.1)$$

gives

$$R_b = \frac{f_{\text{test}}'' R_{\text{test}}}{f''} \quad (20.2)$$

The value of R_b may be determined, therefore, from (20.2) if it is not required that C_s be calculated.

The mid-frequency voltage amplification is determined by (12.2) to be

$$|A|_m = g_m R_b = 9,000 \cdot 10^{-6} \cdot 1,667 = 15$$

Before the value of the compensating inductance can be determined, it is necessary to decide which value of N , Sec. 18, is required. If uniform frequency response up to the upper half-power frequency is desired, N must be 0.5, and from (18.5)

$$L_b = \frac{NR_b}{\omega''} = \frac{0.5 \cdot 1,667}{2\pi \cdot 3 \cdot 10^6} = 44.2 \mu\text{h}$$

For constant time delay, however, N should be 0.32, and

$$L_b = \frac{0.32 \cdot 1,667}{2\pi \cdot 3 \cdot 10^6} = 28.3 \mu\text{h}$$

Most applications, on the other hand, would require that both frequency response and time delay be considered. The value of $N = 0.44$ would be required for this compromise case, and

$$L_b = \frac{0.44 \cdot 1,667}{2\pi \cdot 3 \cdot 10^6} = 38.9 \mu\text{h}$$

In practice, the value of L_b is adjusted experimentally to give the desired frequency response and phase relations.

21. Low-frequency Compensation.¹—The voltage amplification and relative phase angle at low frequencies may be controlled by the use of the compensating networks shown in Figs. 17.1*d* and *e*, for example. In each case, the network R_b and C_b is a part of the usual decoupling network utilized in a multistage amplifier in which the plate-polarizing potential for each tube is obtained from a common power supply. Since the mathematical analysis of the circuit becomes quite involved, only the final results for each case are presented.

When fixed biasing is used and the low-frequency distortion is not caused by cathode degeneration, the circuit of Fig. 17.1*d* maintains the voltage amplification at low frequencies approximately equal to the mid-frequency voltage amplification and gives a minimum relative phase angle when

$$R_c C = \frac{R_b R_b'}{R_b + R_b'} C_b \quad (21.1)$$

or if

$$R_b' \gg R_b$$

when

$$R_c C = R_b C_b \quad (21.2)$$

it being assumed that the a-c plate resistance of the tube is large when compared with the plate-load impedance, which is the case for broad-band amplifiers employing pentodes. For these conditions, the voltage amplification at low frequencies becomes

$$|A_L| \doteq \frac{g_m R_b}{1 - j \frac{f' R_b}{f R_b'}} \quad (21.3)$$

where f' is the lower half-power frequency before compensation, and f is the frequency of the applied signal. When $f = f'$ and

¹ R. L. FREEMAN and D. D. SCHANTZ, Video Amplifier Design, *Electronics*, **10**, 8, 22, August, 1937; D. L. JAFFE, Wide Band Amplifiers and Frequency Multiplication, *Electronics*, **15**, 4, 56, April, 1942.

$R_b'/R_b \gg 1$, (21.3) reduces to

$$|A_L| \doteq g_m R_b$$

the mid-frequency value.

The relative phase angle of the compensated stage at low frequencies is given by

$$\tan \theta \doteq \frac{f' R_b}{f R_b'} \quad (21.4)$$

and the tangent of the relative phase angle is reduced in proportion to the ratio of R_b' to R_b .

When cathode biasing is employed and the low-frequency distortion is caused by cathode degeneration only, Fig. 17.1e, satisfactory low-frequency voltage amplification and relative phase angle are obtained if

$$\left. \begin{aligned} R_b' &= (g_m R_b) R_k \\ C_k &= (g_m R_b) C_b \end{aligned} \right\} \quad (21.5)$$

Under these conditions the time constants of the cathode-biasing circuit and the decoupling or compensating circuit are equal.

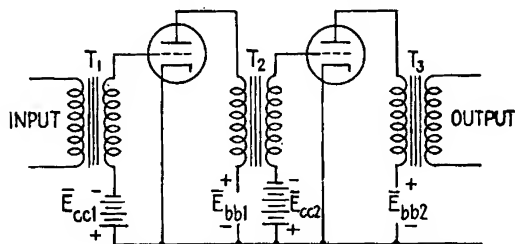


FIG. 22.1.—Two-stage audio-frequency transformer-coupled amplifier.

22. Untuned Transformer-coupled Amplifiers.—The schematic circuit diagram for a two-stage transformer-coupled audio-frequency amplifier is shown in Fig. 22.1. Transformer T_1 is an input transformer; T_2 is an interstage transformer; T_3 is an output transformer. The input transformer is used primarily as a matching device between a low-impedance source and the high input impedance of the tube. It serves also to isolate the grid circuit of the input stage from all steady potentials that may exist in the source and permits the use of an ungrounded source with a grounded amplifier. When the input potential is obtained from a balanced line, the primary of the input transformer may be center-tapped. The interstage transformer serves as the coupling network between

the stages and isolates the steady components of voltage and current in the plate circuit of the preceding tube from the grid circuit of the following tube. The output transformer couples the power-amplifier stage to the load. The impedance ratio of the transformer is chosen so that the plate-load impedance presented to the tube is that required for proper operation. Output transformers often have tapped secondaries to match loads of different impedances.

The advantages of using an interstage transformer rather than resistance coupling are the increase in voltage amplification due to the step-up ratio of the transformer, the low resistance of the primary winding, and the adaptability of the transformer with center-tapped secondary to push-pull amplifier circuits.

It has been shown that the maximum voltage amplification obtainable with either a resistance-coupled or an impedance-coupled amplifier is the μ of the tube used. It is possible, over a limited frequency range, to obtain a voltage amplification of $\mu N_2/N_1$, where N_2 and N_1 are the number of turns on the secondary and primary, when a well-designed transformer is used. At the same time, the limitations on transformer construction require the use of a relatively low- μ triode for satisfactory frequency response. In high-quality audio-frequency amplifiers it is possible, therefore, to obtain a larger voltage amplification with a resistance-coupled pentode stage than with a transformer-coupled triode stage.

On account of the relatively low d-c resistance of the primary winding of the interstage transformer, the required plate power-supply voltage \bar{E}_{bb} need not be much greater than the plate-polarizing potential \bar{E}_{b0} specified for the tube. This is of considerable importance when tubes having large quiescent plate currents are employed.

The use of push-pull input transformers is discussed in Sec. 24.

The important disadvantages of the use of interstage transformers are the poor frequency response, the relatively narrow frequency range over which the voltage amplification is constant, magnetic coupling to other stages of the amplifier and to sources of hum and noise, and also cost, weight, and size.

The frequency response of a transformer-coupled audio-frequency amplifier may be controlled by methods discussed in this section, but in no case is it possible to obtain band widths comparable with those obtainable with resistance coupling. The actual and equivalent circuit diagrams of a transformer-coupled stage are shown in Fig. 22.2.

The circuit of Fig. 22.2b is the lumped-element equivalent of the actual circuit of Fig. 22.2a.¹ In the equivalent circuit, the transformer is represented by an ideal transformer of turns ratio $\alpha = N_2/N_1$, combined with lumped circuit elements to account for the primary and secondary winding resistances, the primary and secondary leakage inductances, the core losses, the primary,

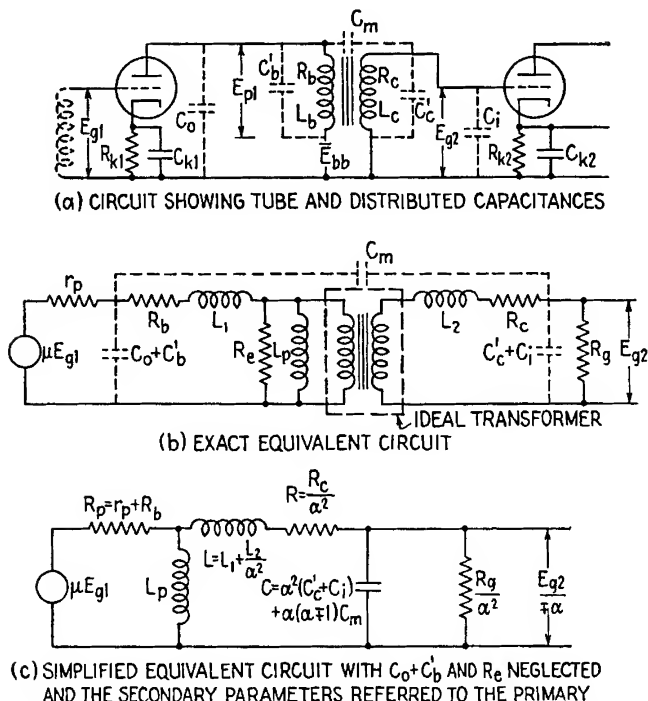


FIG. 22.2.—Untuned transformer-coupled amplifier stage and equivalent circuits.

secondary, and interwinding capacitances, and the incremental inductance of the primary winding. The inductance L_p is the difference between the incremental primary inductance and the primary leakage inductance L_1 . In the usual iron-core audio-frequency transformer, however, the coefficient of coupling is very near unity, 0.998 for example, and L_p is approximately equal to the incremental primary inductance.

It has been shown experimentally that the circuit of Fig. 22.2b may be modified to that shown in Fig. 22.2c without introducing

¹ G. KOEHLER, The Design of Transformers for Audio-frequency Amplifiers with Preassigned Characteristics, *Proc. I.R.E.*, **16**, 22, 1742, December, 1928.

appreciable error. This is accomplished by neglecting the core losses (represented by R_c) and the total primary shunting capacitance $C_o + C_b'$, by combining a capacitance that is approximately equivalent to C_m with the secondary shunting capacitance $C_o' + C_i$ and by referring the parameters in the secondary circuit to the primary side of the ideal transformer. The location of the incremental primary inductance in the circuit has been changed so that it is in shunt with the generator. This is justified in practice because L_p is large when compared with L_1 , and at low frequencies the voltage drop across L_1 is negligibly small when compared with the voltage drop across L_p , Fig. 22.2b. At higher frequencies L_p may be neglected.

The total leakage inductance L is small when compared with the incremental primary inductance L_p in all except the poorest transformers. There exist in the circuit, therefore, two distinct resonant frequencies. One occurs at low frequencies corresponding to the parallel-resonant frequency of L_p and C , in which case the reactance of the leakage inductance L may be neglected when compared with the reactance of C . The other corresponds to the series-resonant frequency of L and C ; it occurs at a high frequency for which the reactance of L_p is essentially an open circuit.

It is convenient, therefore, to divide the frequency-response problem of an interstage transformer into the low-frequency, mid-frequency, and high-frequency ranges similar to those used for the resistance-coupled amplifier.

The low-frequency range is defined as that range of frequencies for which the reactance of the equivalent capacitance C is large and the reactance of the leakage inductance is negligibly small. In the low-frequency range, the equivalent circuit of Fig. 22.2c reduces to the circuit of Fig. 22.3a.

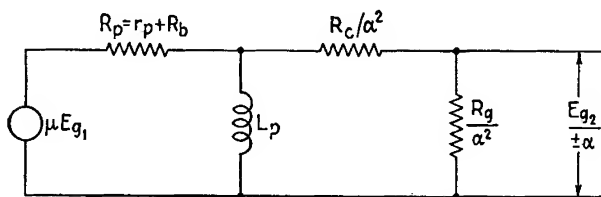
In the mid-frequency range the practical equivalent circuit is that of Fig. 22.3b. In this range of frequencies, the incremental primary inductance L_p is neglected since its reactance is so high that it is essentially an open circuit. Nevertheless, the frequency is not sufficiently high for the leakage inductance L and the equivalent capacitance C to become important.

At higher frequencies L_p is even more negligible; but the reactance of L has appreciable magnitude, and the reactance of the equivalent capacitance becomes small enough to give appreciable shunting effect. Then the equivalent circuit shown in Fig. 22.3c is essentially correct.

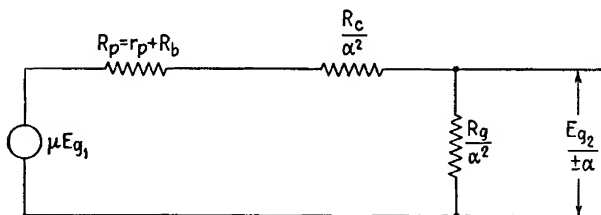
In most practical cases the mid-frequency equivalent circuit is valid for the frequency range determined by

$$\frac{10R_p}{L_p} < \omega < \frac{1}{10\sqrt{LC}} \quad (22.1)$$

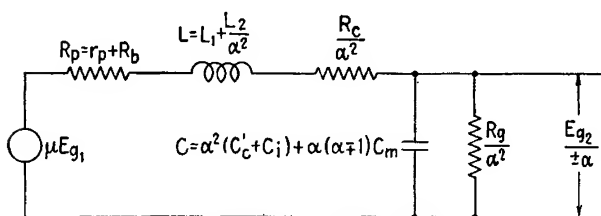
where $R_p = r_p + R_b$.



(a) LOW-FREQUENCY EQUIVALENT CIRCUIT



(b) MID-FREQUENCY EQUIVALENT CIRCUIT



(c) HIGH-FREQUENCY EQUIVALENT CIRCUIT

FIG. 22.3.—Simplified practical equivalent circuits of a transformer-coupled amplifier stage for the low-, mid-, and high-frequency ranges.

At low frequencies for which the equivalent circuit of Fig. 22.3a is valid, a repeated application of the voltage-divider formula yields

$$\frac{E_{g2}}{\pm\alpha} = \frac{R_g}{R_g + R_c} \frac{j\omega L_p \left(\frac{R_c}{\alpha^2} + \frac{R_g}{\alpha^2} \right)}{\frac{R_c}{\alpha^2} + \frac{R_g}{\alpha^2} + j\omega L_p} \frac{\mu E_{g1}}{R_p + j \frac{\omega L_p \left(\frac{R_c}{\alpha^2} + \frac{R_g}{\alpha^2} \right)}{\frac{R_c}{\alpha^2} + \frac{R_g}{\alpha^2} + j\omega L_p}} \quad (22.2)$$

from which

$$A_l = \frac{E_{g2}}{E_{g1}} = \pm \alpha \mu \frac{1}{1 + \frac{R_c}{R_g} + \frac{\alpha^2 R_p}{R_g} - j \frac{R_p}{\omega L_p} \left(1 + \frac{R_c}{R_g}\right)} \quad (22.3)$$

In many applications R_g is the input resistance to a triode that is operated with sufficient negative grid bias to prevent grid current at all times. In all such cases, $R_g \gg R_c$ and $R_g/\alpha^2 \gg R_p$, and the complex voltage amplification becomes

$$A_l = \pm \alpha \mu \frac{1}{1 - j \frac{R_p}{\omega L_p}} \quad (22.4)$$

Equation (22.4) applies at low frequencies for all Class A_1 transformer-coupled amplifiers except those in which a loading resistance is connected across the secondary terminals of the transformer. Equation (22.3) is applicable to circuits where a loading resistor is used. The magnitude of the voltage amplification, from (22.4), is

$$|A_l| = \alpha \mu \frac{1}{\sqrt{1 + (R_p/\omega L_p)^2}} \quad (22.5)$$

and the relative phase angle θ_l is given by

$$\tan \theta_l = + \frac{R_p}{\omega L_p} \quad (22.6)$$

A similar pair of equations for the more general case may be obtained from (22.3).

In the mid-frequency range

$$\frac{E_{g2}}{\pm \alpha} = \frac{R_g}{\alpha^2} \frac{\mu E_{g1}}{R_p + \frac{R_c}{\alpha^2} + \frac{R_g}{\alpha^2}}$$

and

$$A_m = \frac{E_{g2}}{E_{g1}} = \pm \alpha \mu \frac{1}{1 + \frac{R_c}{R_g} + \frac{\alpha^2 R_p}{R_g}} \quad (22.7)$$

which reduces to

$$A_m = \pm \alpha \mu \quad (22.8)$$

when $R_g \gg R_c$ and $R_g/\alpha^2 \gg R_p$. The relative phase angle in the mid-frequency range is approximately zero.

In the high-frequency equivalent circuit, Fig. 22.3c, if

$$Z_1 = R_p + \frac{R_c}{\alpha^2} + j\omega L$$

and

$$Z_2 = \frac{R_g/\alpha^2}{1 + j\omega C \frac{R_g}{\alpha^2}}$$

(Z_2 denoting the impedance of the parallel combination of C and R_g/α^2), then

$$\frac{E_{g2}}{\pm \alpha} = \frac{Z_2}{Z_1 + Z_2} \mu E_{g1} \quad (22.9)$$

Equation (22.9) reduces to

$$A_h = \pm \alpha \mu \frac{1}{1 + \frac{\alpha^2 R_p}{R_g} + \frac{R_c}{R_g} - \omega^2 LC + j \left[\omega C \left(R_p + \frac{R_c}{\alpha^2} \right) + \frac{\omega L \alpha^2}{R_g} \right]} \quad (22.10)$$

when substitutions for Z_1 and Z_2 are made and the expression simplified.

When $R_g/\alpha^2 \gg R_p$, $R_g \gg R_c$, and $R_g/\alpha^2 \gg \omega L$, equation (22.10) reduces to

$$A_h = \pm \alpha \mu \frac{1}{1 - \omega^2 LC + j\omega C \left(R_p + \frac{R_c}{\alpha^2} \right)} \quad (22.11)$$

It follows that

$$|A_h| = \frac{\alpha \mu}{\sqrt{(1 - \omega^2 LC)^2 + \omega^2 C^2 \left(R_p + \frac{R_c}{\alpha^2} \right)^2}} \quad (22.12)$$

and¹

$$\tan \theta_h = - \frac{\omega C \left(R_p + \frac{R_c}{\alpha^2} \right)}{1 - \omega^2 LC} \quad (22.13)$$

¹ Equations (22.12) and (22.13) are often expressed in terms of the Q_r of the circuit. Q_r is defined by

$$Q_r = \frac{\omega_r L}{R_p + \frac{R_c}{\alpha^2}}$$

where $\omega_r = 1/\sqrt{LC}$ is the series-resonant frequency of the leakage inductance and the equivalent capacitance. The magnitudes of the voltage amplification

Equation (22.5) indicates that the low-frequency response of a transformer-coupled amplifier is improved by increasing the primary incremental inductance of the transformer, a constant-

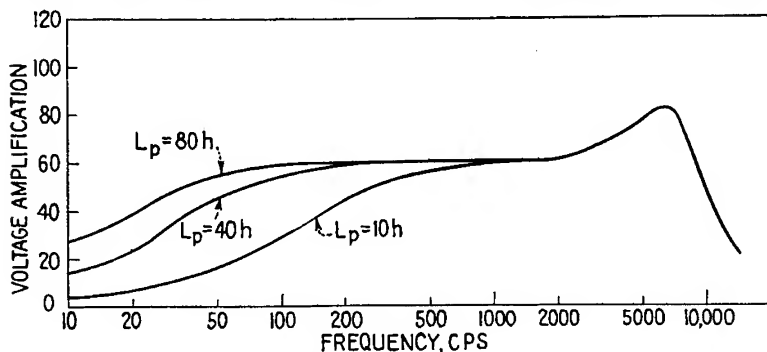


FIG. 22.4.—Variations of $|A|$ as a function of L_p .

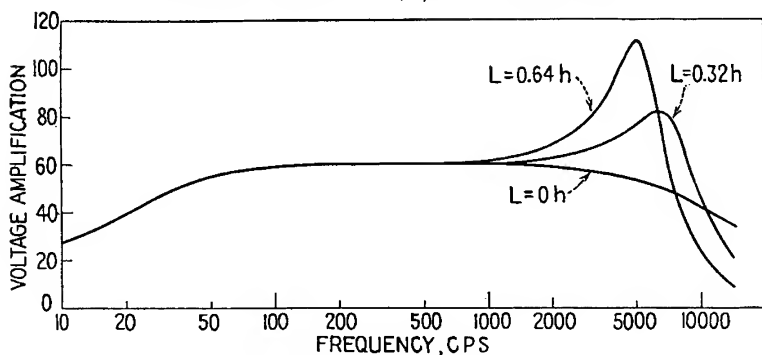


FIG. 22.5.—Effect on frequency response of varying the leakage inductance.

turns ratio and a given R_p being assumed. The variation in the voltage amplification as a function of L_p of a transformer-coupled stage is shown in Fig. 22.4. For good low-frequency response, therefore, it is essential that the number of primary turns be high.

The high-frequency response, (22.12), is dependent upon the leakage inductance L and the equivalent secondary capacitance

and relative phase angle at high frequencies are given by

$$|A_h| = \frac{\alpha\mu}{\sqrt{\left[1 - \left(\frac{\omega}{\omega_r}\right)^2\right]^2 + \left(\frac{\omega}{\omega_r}\right)^2 \frac{1}{Q_r^2}}}$$

and

$$\tan \theta_h = -\frac{1/Q_r}{\frac{\omega_r}{\omega} - \frac{\omega}{\omega_r}}$$

C in addition to R_p and R_e . The effects of varying L and C are shown in Figs. 22.5 and 22.6.

It is seen from Fig. 22.5 that some leakage inductance is desirable in order to compensate for the effect of the equivalent secondary capacitance. On the other hand, by decreasing both the leakage inductance and the equivalent shunting capacitance, the response

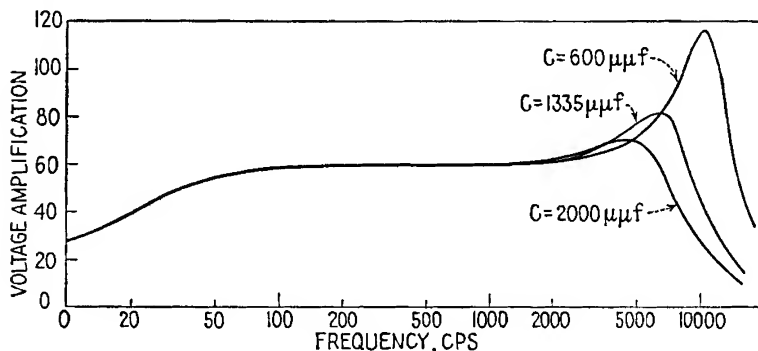


FIG. 22.6.—Effect on frequency response of varying the equivalent secondary capacitance.

of the transformer can be extended to higher frequencies. Since the equivalent secondary capacitance is

$$C = \alpha^2(C_e' + C_i) + \alpha(\alpha \mp 1)C_m \quad (22.14)$$

where C_e' is the distributed capacitance of the secondary, C_i the input capacitance to the second stage, and C_m the interwinding capacitance, the high-frequency response is a function of the interwinding capacitance as well as the actual secondary shunting capacitance $C_e + C_i$.

The curves plotted in Fig. 22.7 illustrate the effect of varying the resistance R_p in the primary circuit of the transformer. These curves were calculated by means of (22.5), (22.8), (22.12), for the transformer constants indicated in the example at the end of this section. Even though the primary-circuit resistance is the sum of the variational plate resistance of the tube and the resistance of the primary winding, the a-c plate resistance of the tube is usually large when compared with R_b . The low-frequency response is improved and the high-frequency response is made less desirable by a small R_p . When the primary-circuit resistance is large, both low and high frequencies are attenuated excessively. The over-all frequency response is improved by means of a medium (or optimum) value of R_p , which prevents the sharp high-frequency peak without

causing serious frequency distortion in the low-frequency range. It is essential, therefore, that an interstage transformer be used with a tube having the variational plate resistance for which the transformer is designed.

Because of the value of R_p (approximately 10,000 ohms for low- μ triodes) and the resulting large number of primary turns required for good low-frequency response, the practical limit of the turns ratio α is approximately *three*. Larger values of α require

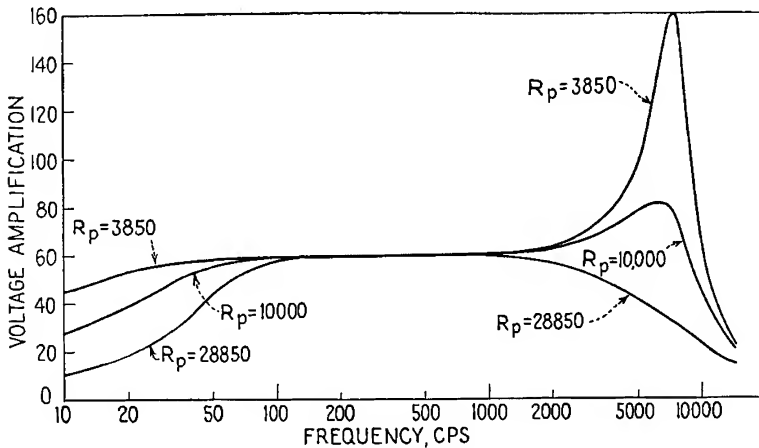


FIG. 22.7.—Effect on frequency response of varying the primary circuit resistance.

proportionately larger values of secondary turns, winding space, and core size, which produce increased leakage inductance and distributed capacitance. It follows, therefore, that an interstage transformer with a large turns ratio has an impaired high-frequency response.

The use of tetrodes and pentodes with interstage transformers necessitates excessive primary turns for satisfactory low-frequency voltage amplification. However, tetrodes and pentodes may be used by placing a resistor in shunt across the primary winding of the transformer. It may be shown by applying Thévenin's theorem that this resistor effectively reduces the primary-circuit resistance and *also* the effective emf of the equivalent generator. Thus, while the low-frequency response is improved, the mid- and high-frequency voltage amplification is reduced by the shunting resistor.

The effect of the resistor R_g may be determined from (22.3), (22.7), (22.10). Curves showing the frequency response of a

transformer-coupled stage as a function of R_g are plotted in Fig. 22.8. The load resistance R_l reduces the voltage amplification at all frequencies. The greatest reduction occurs, however, in the

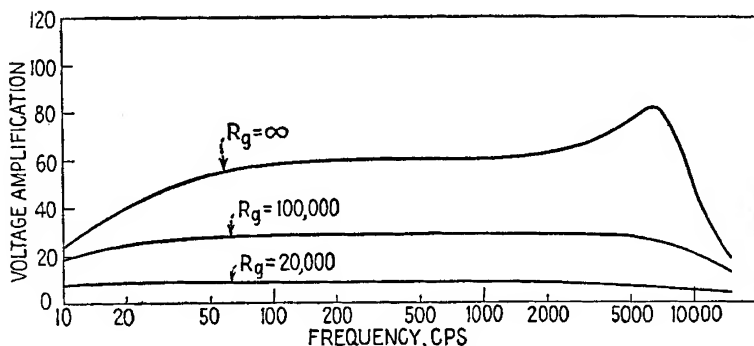


FIG. 22.8.—Effect on frequency response of varying R_g .

vicinity of the high-frequency peak. It is desirable, therefore, to place a resistor in shunt with the secondary winding when uniform frequency response is more important than voltage amplification.

Example.—The constants used in calculating the data for Figs. 22.4 to 22.8 may be considered as typical for interstage transformers. These constants are

Turns ratio $\alpha = 3$

Coefficient of coupling $k = 0.998$

Incremental primary inductance $L_p = 80$ henrys

Primary leakage inductance $L_1 = 0.16$ henry

Secondary leakage inductance $L_2 = 1.44$ henry

Distributed capacitance of primary $C_b' = 15 \mu\text{mf}$

Distributed capacitance of secondary $C_c' = 30 \mu\text{mf}$

Interwinding capacitance $C_m = 87.5 \mu\text{mf}$

A-c resistance of primary $R_b = 1,500$ ohms

A-c resistance of secondary $R_c = 7,500$ ohms

It is assumed that the triode is polarized so that $r_p = 10,000$ ohms, that $\mu = 20$, that the output and input capacitance of the tubes are $C_o = 20 \mu\text{mf}$ and $C_i = 60 \mu\text{mf}$. Therefore, the equivalent circuit elements of Fig. 22.3 are

$$R_p = r_p + R_b = 11,500 \text{ ohms}$$

$$L = L_1 + \frac{L_2}{\alpha^2} = 0.16 + \frac{1.44}{9} = 0.32 \text{ henry}$$

$$C = \alpha^2(C_c' + C_i) + \alpha(\alpha - 1)C_m = 9(30 + 60) + 3(2)87.5 = 1,335 \mu\text{mf}$$

A typical calculation at $f = 32$ cps ($\omega = 200$ radians/sec) for the low-frequency voltage amplification and relative phase angle of the stage, using (22.5) and (22.6), yields

$$|A_i| = \frac{\alpha\mu}{\sqrt{1 + (R_p/\omega L_p)^2}} = \frac{60}{\sqrt{1 + \left(\frac{11,500}{200 \cdot 80}\right)^2}} = \frac{60}{1.23} = 48.7$$

$$\tan \theta_i = + \frac{R_p}{\omega L_p} = 0.719 \quad \text{or} \quad \theta_i = 35^\circ 43'$$

In the high-frequency range the voltage amplification and relative phase angle may be calculated by means of (22.10) and (22.11). At $f = 6,380$ cps ($\omega = 40,000$ radians/sec),

$$|A_h| = \frac{\alpha\mu}{\sqrt{(1 - \omega^2 LC)^2 + \omega^2 C^2 \left(R_p + \frac{R_c}{\alpha^2}\right)^2}} = \frac{60}{\sqrt{(1 - 0.684)^2 + 0.434}}$$

$$|A_h| = \frac{60}{\sqrt{0.534}} = \frac{60}{0.731} = 82.1$$

$$\tan \theta_h = - \frac{\omega C \left(R_p + \frac{R_c}{\alpha^2}\right)}{1 - \omega^2 LC} = - \frac{0.659}{0.316} = -2.08$$

$$\theta_h = -64^\circ 22'$$

23. Single-tube Class A Audio-frequency Power Amplifiers.—

The purpose of the output stage of an a-f amplifier is to supply the power required to operate a loudspeaker, cutting head, or similar device. When employed in a transmitter, the output stage of the audio-frequency amplifier supplies the modulating power. It is desired, in general, to design the output stage of an amplifier for the largest power output and plate-circuit efficiency that are consistent with the maximum plate dissipation of the tube and the distortion tolerances allowed. The varying component of the total voltage applied to the grid of a power stage is often of such amplitude that the operation of the tube is not limited to the linear portion of its characteristics. Harmonic, or nonlinear, distortion is introduced by the large curvature of the characteristic curves of the tube at low plate currents and by the distortion of the input signal wave when its amplitude exceeds the grid-polarizing potential, thus causing grid current to flow. Grid-circuit distortion is readily prevented by not permitting grid current to flow. Nonlinear distortion is minimized by the use of the proper plate-load resistance and by not permitting the instantaneous plate current to fall in the region of large curvature. Nonlinear distortion may be made small, therefore, by operating the tube Class A, which is characterized, however, by relatively low power output and plate-circuit efficiency.

The dynamic characteristics of a typical power triode for three values of a-c plate-load resistance \bar{R}_b have been superimposed

upon the static characteristics of the tube, Fig. 23.1. The quiescent point Q , determined by \bar{E}_{b0} and \bar{E}_{c0} , has been chosen for typical Class A operation as recommended by the tube manufacturer. In Fig. 23.1a, \bar{E}_{c0} is midway between $e_c = 0$ and $e_{c \min}$ when \tilde{R}_b is 2,500 ohms. A method of locating the optimum Q point is discussed later in this section. The dynamic characteristics, which combine the tube characteristics with the effect of the plate-load resistance, are more nearly linear than the static characteristics of the tube.

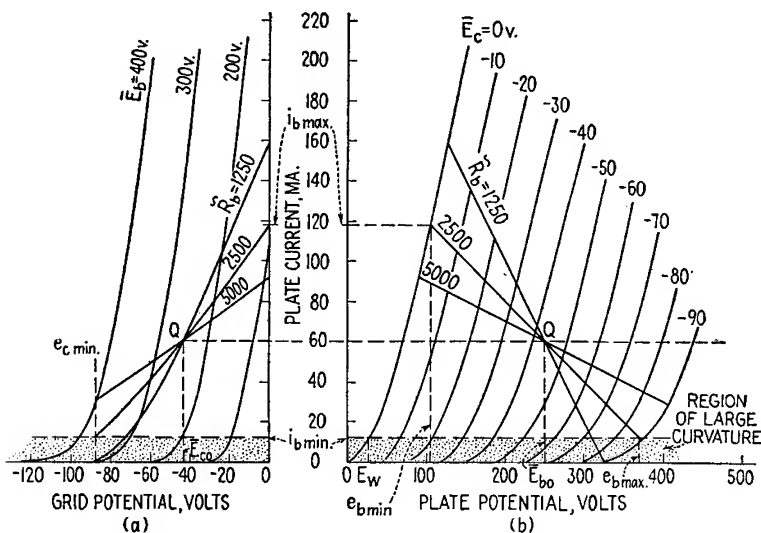


FIG. 23.1.—A typical power triode in Class A₁ operation.

Furthermore, the larger \tilde{R}_b , the more nearly linear the dynamic characteristic.

For values of \tilde{R}_b equal to and greater than the optimum value \tilde{R}_{b0} , 2,500 ohms in this illustration, the maximum amplitude of grid voltage is limited by the condition that the grid current be zero at all times. This restriction requires that \bar{E}_g never exceed \bar{E}_{c0} . It is possible that grid current will flow when the total instantaneous grid voltage is slightly negative, thus limiting still further the maximum amplitude of the grid voltage. It is reasonable in most cases, however, to assume that no appreciable grid current flows until the grid potential is positive with respect to the cathode. On the other hand, when \tilde{R}_b is less than \tilde{R}_{b0} , the maximum value of \bar{E}_g is limited by the condition that the path of operation shall not enter the region of large curvature; for example, the plate

current must not become less than $i_{b \min}$, Fig. 23.1. The minimum permissible value of plate current is determined by the maximum amount of distortion specified.

Since the d-c resistance of the plate-load impedance (usually an output transformer) is relatively small in output stages, the quiescent plate voltage is very nearly equal to the power-supply potential.

Optimum power output is defined as the maximum power output that can be obtained from a given tube with a given quiescent value of plate potential \bar{E}_{b0} and with a specified maximum harmonic (nonlinear) distortion. The plate-load resistance for which optimum power output is available is called *optimum load resistance*. The grid bias corresponding to the conditions of optimum load resistance is defined as *optimum bias*.

Even though the conditions for optimum operation of a tube can be determined accurately only by either graphical or laboratory means, the theoretical values of optimum load resistance, optimum power output, and optimum bias may be determined by assuming that the static characteristics are linear, parallel, and equally spaced for equal increments of the grid voltage over the operating range. It can be shown that, for the idealized case of linear operation of triodes with specified \bar{E}_{b0} , optimum power output is obtained when

$$R_L = 2r_p \quad (23.1)$$

With this optimum value of plate-load resistance, the optimum power output¹ is

$$(\bar{P}_L)_{\text{opt}} = \frac{2}{3} \mu g_m |E_g|^2 \quad (23.2)$$

Since the amplitude of the varying component of the grid voltage \hat{E}_g is approximately equal to the quiescent grid voltage \bar{E}_{c0} , (23.2) becomes

$$(\bar{P}_L)_{\text{opt}} \doteq \frac{1}{3} \mu g_m \bar{E}_{c0}^2 \quad (23.3)$$

It can be shown by similar considerations that, if $i_{b \min}$ is held constant as \bar{R}_L is varied, the optimum grid bias is given approximately by

$$(\bar{E}_c)_{\text{opt}} = -\frac{3}{4} \frac{\bar{E}_{b0} - \bar{E}_w}{\mu} \quad (23.4)$$

where \bar{E}_w is the plate potential for which the plate current is $i_{b \min}$

¹ The maximum power output with 10 per cent total harmonic distortion is sometimes called the "maximum undistorted power output."

with zero grid voltage, Fig. 23.1*b*. If $i_{b \min}$ is held constant as \tilde{R}_L is varied or if the ratio of $i_{b \max}$ to $i_{b \min}$ is held constant,¹ the optimum grid bias is given by

$$(\bar{E}_c)_{\text{opt}} = \frac{3 \frac{i_{b \max}}{i_{b \min}} - 1}{4 \frac{i_{b \max}}{i_{b \min}}} \frac{\bar{E}_{b0}}{\mu} \quad (23.5)$$

It is shown in Chap. V, Sec. 5, that maximum power is delivered to a resistive load by a constant-voltage generator when

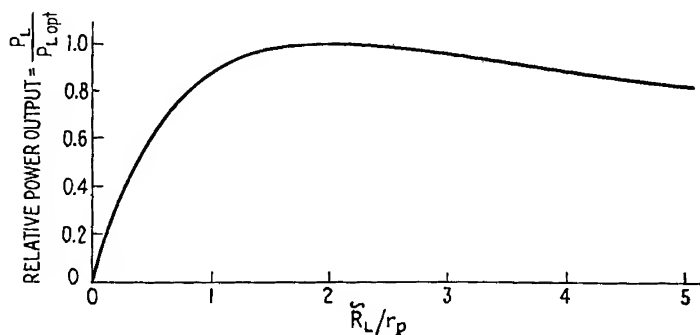


FIG. 23.2.—Theoretical relative power output from a triode in Class A_1 linear operation as a function of \tilde{R}_L/r_p .

the load resistance is equal to the generator resistance. Even though it might appear that (23.1) contradicts this well-established principle, this is not the case. A larger amplitude of the varying component of grid voltage can be used when \tilde{R}_L equals $2r_p$ than when \tilde{R}_L equals r_p for a given \bar{E}_{b0} and specified maximum distortion with Class A_1 operation. The maximum power output is larger, therefore, when \tilde{R}_L is equal to $2r_p$. The variation in the relative power output as a function of the ratio of \tilde{R}_L to r_p is illustrated in Fig. 23.2. The maximum is so broad that the power output is at least 88 per cent of the optimum value for values of \tilde{R}_L between r_p and $4r_p$.

It is rarely possible to obtain the theoretical optimum conditions of operation, for the idealized tube characteristics are seldom

¹ Equations (23.1) to (23.5) are derived in many standard texts, *e.g.*, H. J. Reich, "Theory and Application of Electron Tubes," pp. 249–259, McGraw-Hill Book Company, Inc., 1939; R. S. Glasgow, "Principles of Radio Engineering," pp. 179–183, McGraw-Hill Book Company, Inc., 1936; F. E. Terman, "Radio Engineering," pp. 277–282, McGraw-Hill Book Company, Inc., 1937; J. Millman and S. Seely, "Electronics," pp. 633–637, McGraw-Hill Book Company, Inc., 1941.

realized. It is desirable, therefore, to resort to graphical means in most practical applications.

The a-c power delivered to a nonreactive load may be calculated from the relationship

$$\bar{P}_L = \frac{(e_{b \max} - e_{b \min})(i_{b \max} - i_{b \min})}{8} \quad (23.6)$$

since the a-c current and voltage are in phase and $\cos \theta$ is unity. Thus, from Fig. 23.1a, when \bar{R}_L is 2,500 ohms, $i_{b \max}$ is 118 ma, $i_{b \min}$ is 12 ma, $e_{b \max}$ is 368 volts, $e_{b \min}$ is 105 volts, and the a-c

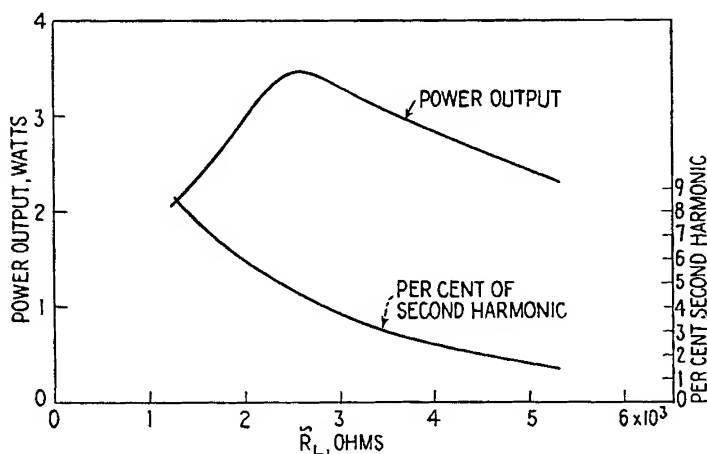


FIG. 23.3.—Power output and per cent second harmonic content for the triode illustrated in Fig. 23.1.

power delivered to the load is 3.5 watts. The a-c power output for the conditions of operation illustrated in Fig. 23.1 is plotted as a function of \bar{R}_L in Fig. 23.3. Optimum power output is obtained when \bar{R}_L is 2,500 ohms. This is approximately three times the variational plate resistance r_p of the triode. Thus, for a given tube operated with specified \bar{E}_{b0} and maximum harmonic distortion, it is essential that the optimum bias and optimum load resistance be determined graphically by trial and error. If \bar{E}_{b0} is not specified, this must be determined first. Since the power output is proportional to \bar{E}_{b0}^2 , the largest permissible value of \bar{E}_{b0} is chosen. The maximum value of \bar{E}_{b0} is limited by the maximum allowable plate dissipation of the tube and by the available power supply. For small power tubes the latter limitation is often of importance because of the cost of high-voltage rectifiers. The choice of the

Q point must be such that $\bar{I}_{b0}\bar{E}_{b0}$ does not exceed the maximum allowable plate dissipation of the tube. As a first approximation, \bar{E}_{c0} may be determined from

$$\bar{E}_{c0} \doteq -0.7 \frac{\bar{E}_{b0}}{\mu} \quad (23.7)$$

If $\bar{I}_{b0}\bar{E}_{b0}$ exceeds the maximum allowable plate dissipation, \bar{I}_{b0} is reduced by increasing \bar{E}_{c0} . The optimum load resistance for the specified distortion then is determined graphically by trial and error.

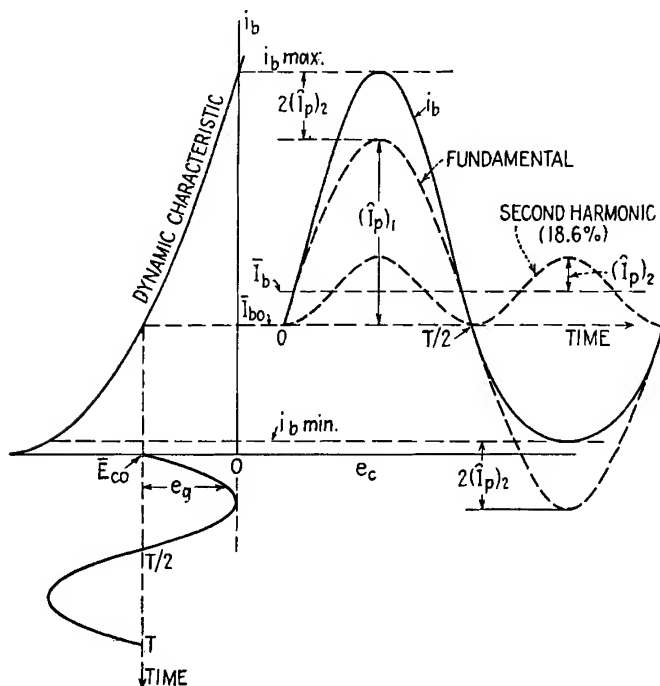


FIG. 23.4.—Nonlinear distortion in a triode.

The nonlinear distortion that occurs in single-tube Class A_1 triode amplifiers consists almost entirely of second harmonic. The plate current of a triode with a sinusoidal voltage applied to its grid is illustrated in Fig. 23.4 together with its fundamental and second harmonic. The amplitudes of the fundamental and second harmonic are $(\hat{I}_p)_1$ and $(\hat{I}_p)_2$. From the figure

$$i_{b \max} = \bar{I}_{b0} + (\hat{I}_p)_1 + 2(\hat{I}_p)_2 \quad (23.8)$$

and

$$i_{b \min} = \bar{I}_{b0} - (\hat{I}_p)_1 + 2(\hat{I}_p)_2 \quad (23.9)$$

Therefore,

$$(\hat{I}_p)_1 = \frac{i_{b \max} - i_{b \min}}{2} \quad (23.10)$$

and

$$(\hat{I}_p)_2 = \frac{i_{b \max} + i_{b \min} - 2\bar{I}_{b0}}{4} \quad (23.11)$$

It follows from (23.10) and (23.11) that the

$$\begin{aligned} \text{Per cent 2d harmonic} &= 100 \frac{(\hat{I}_p)_2}{(\hat{I}_p)_1} \\ &= 100 \frac{i_{b \max} + i_{b \min} - 2\bar{I}_{b0}}{2(i_{b \max} - i_{b \min})} \quad (23.12) \end{aligned}$$

The per cent second harmonic is plotted as a function of \bar{R}_L in Fig. 23.3 for the example of Fig. 23.1. Increasing \bar{R}_L from 2,500 ohms to 3,000 ohms decreases the output power by approximately 6 per cent (from 3.5 to 3.3 watts) while reducing the second harmonic by approximately 20 per cent (from 4.7 to 3.8 per cent).

Equation (23.12) can be written in the form

$$\text{Per cent 2d harmonic} = 100 \frac{\left(\frac{i_{b \max} - \bar{I}_{b0}}{\bar{I}_{b0} - i_{b \min}} \right) - 1}{2 \left(\frac{i_{b \max} - \bar{I}_{b0}}{\bar{I}_{b0} - i_{b \min}} \right) + 1} \quad (23.13)$$

which shows that the per cent second harmonic increases with the ratio of $i_{b \max} - \bar{I}_{b0}$ to $\bar{I}_{b0} - i_{b \min}$. It follows, therefore, that the second harmonic content in a triode increases as the Q point is adjusted for smaller values of \bar{I}_{b0} when \bar{R}_L is constant and diminishes as \bar{R}_L is increased with the Q point fixed. With increased values of \bar{R}_L , the grid bias may be increased, thus permitting larger amplitudes of grid voltage and greater power output without exceeding the specified maximum distortion. Furthermore, the quiescent plate current is reduced, and the plate-circuit efficiency of the tube is increased.

The use of pentodes rather than triodes as Class A_1 power amplifiers results in increased power sensitivity, power output, and plate-circuit efficiency. The power sensitivity of a tube is defined as the ratio of the power output to the square of the rms value of grid-signal voltage. Since μ for pentodes is much larger than for triodes having a g_m of the same order of magnitude,

pentodes require smaller grid-signal voltage for a given power output. This results in increased power sensitivity.

The i_b - e_b characteristics of a typical power pentode are shown in Fig. 23.5b. The minimum value of e_b , corresponding to $i_{b \max}$ for a given \tilde{R}_L , is smaller than could be obtained in a triode without driving the grid positive. Since the resulting amplitude of the varying component of the plate voltage is large when compared

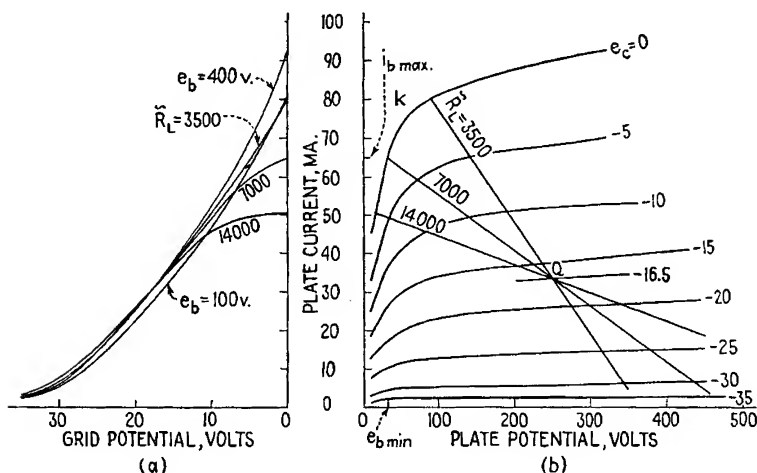


FIG. 23.5.—A typical power pentode in Class A₁ operation.

with that obtainable with a triode of similar plate dissipation and with comparable power-supply voltage, the power output of the pentode for a specified maximum harmonic distortion is greater than that of a similar triode.

The plate current in a pentode is very nearly independent of the plate voltage, except at very low values of e_b , because of the high positive potential and the shielding action of the screen grid. Large values of plate current are obtained in a pentode, therefore, when the plate voltage is relatively low. For this reason the plate-circuit efficiency of power pentodes is greater than for triodes.

The dynamic characteristics of a typical power pentode for three values of \tilde{R}_L , and with the Q point of Fig. 23.5b, have been superimposed upon its static characteristics in Fig. 23.5a. The dynamic characteristic has much greater curvature than for the corresponding characteristic of a triode. Furthermore, the curvature becomes more objectionable when \tilde{R}_L is increased. The power

output may be determined approximately by (23.6), but a more accurate analysis is required before the power output at the fundamental frequency is known. The curvature of the dynamic characteristic of pentodes produces both second and third harmonic components in the plate current. The amplitudes of the fundamental and harmonics can be determined by harmonic analysis either from the i_b - e_b characteristics for a given path of operation or from the corresponding dynamic characteristic. The method¹ presented here is a five-point analysis based upon the instantaneous values of the plate current corresponding to values of e_g equal to 0, $\pm \hat{E}_g/2$, and $\pm \hat{E}_g$. It is assumed that e_g is sinusoidal. For use in equations to follow, the total instantaneous plate currents i_b' and i_b'' are defined as those occurring when e_g is $+\hat{E}_g/2$ and $-\hat{E}_g/2$, respectively; $i_{b \max}$ and $i_{b \min}$ are the maximum and minimum values of plate current; and \bar{I}_{b0} is the quiescent value of plate current. When the fourth and higher harmonics can be neglected,

$$\bar{I}_b = \bar{I}_{b0} + (\hat{I}_p)_2 \quad (23.14)$$

$$(\hat{I}_p)_1 = \frac{i_{b \max} - i_{b \min} + i_b' - i_b''}{3} \quad (23.15)$$

$$(\hat{I}_p)_2 = \frac{i_{b \max} + i_{b \min} - 2\bar{I}_{b0}}{4} \quad (23.16)$$

$$(\hat{I}_p)_3 = \frac{i_{b \max} - i_{b \min} - 2(i_b' - i_b'')}{6} \quad (23.17)$$

When it is not feasible to neglect the fourth and higher harmonics, an analysis employing a larger number of reference points must be employed.

The power output at the fundamental frequency is

$$\bar{P}_L = \frac{(\hat{I}_p)_1^2}{2} \bar{R}_L = \frac{(i_{b \max} - i_{b \min} + i_b' - i_b'')^2}{18} \bar{R}_L \quad (23.18)$$

The power output and harmonic content for the pentode whose characteristics are given in Fig. 23.5 are shown as a function of \bar{R}_L in Fig. 23.6. When $i_{b \max} + i_{b \min}$ is equal to $2\bar{I}_{b0}$, the amplitude of the second harmonic is zero, from (23.16). This occurs in this illustration when \bar{R}_L is 7,300 ohms. It is possible, therefore, to select \bar{R}_L so that there is no second harmonic. This value of \bar{R}_L is much less than r_p and far removed from the value of \bar{R}_L for

¹ E. L. CHAFFEE, A Simplified Harmonic Analysis, *Rev. Sci. Instruments*, **7**, 384, 1936.

maximum power output. For pentodes, the optimum value of \tilde{R}_L is *not* $2r_p$, being more nearly $r_p/10$.

The optimum load resistance for pentode output stages is determined more often by distortion limitations than by power-output considerations. A first approximation for the optimum load resistance for a given pentode operated Class A_1 at a given

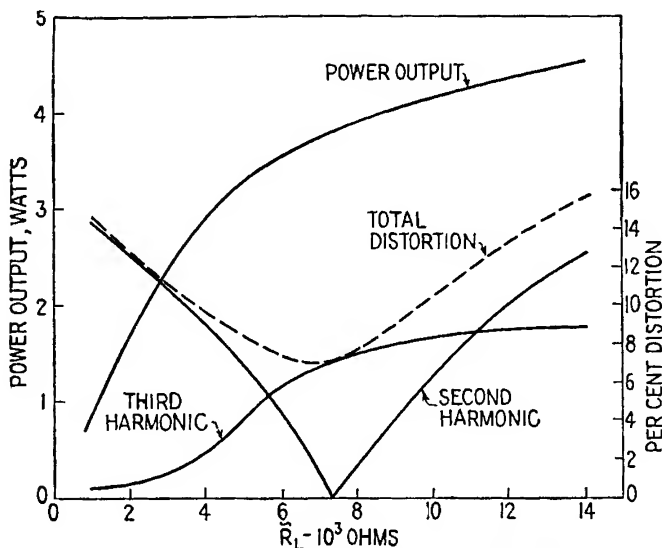


FIG. 23.6.—Power output of the pentode of Fig. 23.5.

Q point may be determined from i_b - e_b characteristics such as are illustrated in Fig. 23.5. The quiescent grid potential determines the maximum amplitude of grid-signal voltage and, in turn, permits the determination of $i_{b \min}$, which is approximately independent of \tilde{R}_L . Since \bar{I}_{b0} and $i_{b \min}$ are known, $i_{b \max}$ for no second harmonic is calculated from (23.16). The minimum value of plate potential, corresponding to $i_{b \max}$, is read from the curve $e_c = 0$. The plate-load resistance for no second harmonic is calculated by

$$\tilde{R}_L = \frac{\bar{E}_{b0} - e_{b \min}}{i_{b \max} - \bar{I}_{b0}} \quad (23.19)$$

and the power output is determined by (23.6) if the power contributed by the harmonics is negligible.¹ The amplitude of the third

¹ When the power contributed by the harmonics is not negligible,

$$P_L = [(\bar{I}_p)_1^2 + (\bar{I}_p)_2^2 + (\bar{I}_p)_3^2 + \cdots] \frac{\tilde{R}_L}{2}$$

harmonic is calculated by (23.17). By determining the power output and harmonic content for several values of \tilde{R}_L slightly less than and slightly greater than the value for no second harmonic, curves similar to Fig. 23.6 may be plotted. \tilde{R}_L is selected for the largest power output that is consistent with the distortion limitations.

In case \tilde{E}_{bb} is specified, it being assumed that $\tilde{E}_{b0} = \tilde{E}_{bb}$, the optimum bias may be selected by determining \tilde{R}_L for no second harmonic for several values of \tilde{E}_c . The value of \tilde{E}_c for which maximum power output is obtained is selected as the optimum bias.

If neither \tilde{E}_{bb} nor \tilde{E}_c is given, the largest plate-polarizing potential consistent with the maximum allowable plate dissipation and optimum power output is selected to a first approximation by assuming that \tilde{I}_{b0} equals $i_{b \max}/2$, where $i_{b \max}$ is the plate current at the knee k of the $e_c = 0$ characteristic, Fig. 23.5. \tilde{E}_b then is chosen so that $\tilde{E}_b (i_{b \max}/2)$ does not exceed the maximum allowable plate dissipation of the tube. \tilde{E}_c and \tilde{R}_L then are determined by the methods previously outlined.

When second harmonic is present in the output, the average value of plate current replaces the quiescent value in the application of the preceding method.

Since these graphical determinations are laborious and time-consuming and the characteristics of a given tube may vary appreciably from the average characteristics provided by the manufacturer, it is preferable in many applications to determine experimentally the optimum conditions of operation.

The i_b - e_b characteristics of beam-power tubes are more nearly straight and parallel for low values of e_b than are those of conventional pentodes. For this reason, greater power output for a given plate-polarizing potential is available without excessive distortion than from pentodes. For minimum distortion, beam-power tubes are usually operated in push-pull.

In power pentodes and beam-power tubes the screen-grid current increases appreciably when the plate potential is small. It is important, therefore, that precautions be taken to limit the screen-grid current so that the maximum allowable screen-grid dissipation is not exceeded. The average screen-grid current \tilde{I}_d may be determined graphically from the i_d - e_b characteristics of the tube by the same methods employed in determining the average plate current. The average power delivered to the screen is given by $\tilde{I}_d \tilde{E}_d$.

The larger the value of \tilde{R}_L , the smaller the plate potential

corresponding to $i_{b\max}$ and the larger the average value of the screen current. Screen dissipation may limit the maximum plate-load resistance. In order to avoid excessive screen dissipation, the load resistance should not be removed from the secondary of the output transformer nor should the plate voltage be removed before the screen voltage.

Since the resistance of the load to which power is delivered by a power amplifier is seldom equal to the optimum load resistance of the tube, output transformers are used to “match” the load to the tube. The transformer characteristics should be such that the input impedance of the loaded transformer is the optimum load resistance for the tube. This condition is satisfied for an ideal transformer when the turns ratio α is such that

$$(\tilde{R}_L)_{\text{opt}} = \frac{\tilde{R}_L}{\alpha^2} \quad (23.20)$$

where \tilde{R}_L is the secondary load resistance and $(\tilde{R}_L)_{\text{opt}}$ is the optimum load resistance for the tube. Output transformers may be considered as ideal transformers in the mid-frequency range, *i.e.*, in the neighborhood of 1,000 cps. At frequencies below and above the mid-frequency range, the transformer characteristics introduce frequency distortion, which can be diminished by use of negative feedback, Sec. 29, or by compensating networks. With pentodes and beam-power tubes it is sometimes desirable to place either a resistance or a resistance in series with a capacitor in parallel with the primary of the output transformer to minimize the effects of the variation of the secondary load impedance (*viz.*, the voice coil of a loudspeaker) with frequency.

24. Push-pull Amplifiers.—The push-pull circuit shown in Fig. 24.1 has several advantages over single-sided circuits for power-output stages. Probably the most important of these is the reduction of nonlinear distortion, which makes it possible to obtain increased power output (compared with single-sided operation) for a specified maximum distortion. Other advantages include, for Class A operation, reduced sensitivity to ripple voltage in the plate power supply, reduced magnetization of the output-transformer core, and less stringent requirements on the cathode by-pass capacitor.

In the circuit of Fig. 24.1, the tubes are assumed to have identical electrical characteristics, and both the secondary of the input

transformer and the primary of the output transformer are assumed to be electrically center-tapped.

When the input signal voltage e_s is zero, the grid bias on each tube is $\bar{E}_{c0} = -\bar{E}_{cc}$ and the plate-polarizing potential on each tube is $\bar{E}_{b0} = \bar{E}_{bb}$, the d-c potential across the halves of the primary of the output transformer being neglected. Thus, the two tubes have identical quiescent plate currents \bar{I}_{b0} , and the total current through the power supply is $2\bar{I}_{b0}$. The magnetomotive force (mmf) in the upper half of the primary of the output transformer is equal

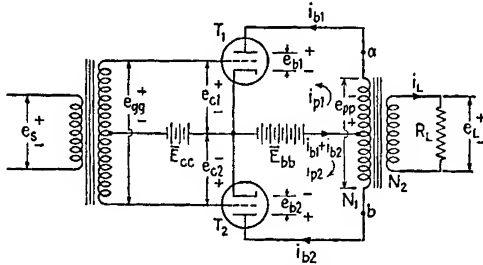


FIG. 24.1.—Basic push-pull circuit with transformer input and output circuits and opposite to that in the lower half, and the net magnetization of the core is zero. This cancellation of the magnetomotive forces in the primary of the output transformer permits push-pull output transformers to be designed with smaller iron cores than would be required for single-sided operation with quiescent plate currents of the same order of magnitude. For the same reason, push-pull operation is often employed in voltage amplifiers in which coupling is made by miniature transformers.

Variations in \bar{E}_{bb} , such as a ripple voltage in the power supply and variations in the power-supply voltage caused by poor regulation, are applied simultaneously to the two tubes, producing equal increments in plate currents. Since these increments of current flow in opposite directions in the halves of the primary of the output transformer, the total change in magnetic flux due to them is zero and no voltage is induced in the secondary circuit. With push-pull operation, therefore, it is not necessary to smooth the rectifier output as thoroughly as for single-sided stages.

Assume the tubes of Fig. 24.1 to be biased for Class A operation and the amplitudes of the varying components of the grid voltages to be so small that the operation of each tube is limited to the linear region of the tube characteristics. The signal voltage having the instantaneous value e_s , with the polarity indicated, induces e_{pp} in

the secondary, causing the grid potentials of T_1 and T_2 to vary with equal magnitudes and opposite signs. The resulting variations in plate currents are of equal magnitude, i_{b1} increasing as i_{b2} decreases, and vice versa. The varying components of the plate currents have the conventional directions indicated by the arrows labeled i_{p1} and i_{p2} . Then

$$i_{b1} = \bar{I}_{b0} + i_{p1} \quad (24.1)$$

$$i_{b2} = \bar{I}_{b0} + i_{p2} \quad (24.2)$$

Since the varying components have equal magnitudes and opposite signs,

$$i_{p2} = -i_{p1} \quad (24.3)$$

The current through the plate power supply is the sum of the two plate currents,

$$i_{bb} = i_{b1} + i_{b2} \quad (24.4)$$

which from (24.1) to (24.3) is equal to $2\bar{I}_{b0}$ and contains no varying component. Regulation of the power-supply potential, therefore, is of secondary importance when the tubes of a push-pull combination are restricted to linear Class A operation. Furthermore, the absence of a varying component of current through the power supply permits the use of a cathode-bias resistor without a by-pass capacitor.

Since the sum of the instantaneous magnetomotive forces around the core in a given direction is zero in an ideal transformer,

$$\frac{N_1}{2} i_{b1} - \frac{N_1}{2} i_{b2} - i_L N_2 = 0 \quad (24.5)$$

where N_1 is the total number of primary turns and N_2 is the total number of secondary turns. Then

$$i_L = \frac{N_1}{N_2} \frac{i_{b1} - i_{b2}}{2} \quad (24.6)$$

From (24.1) to (24.3),

$$i_L = \frac{N_1}{N_2} i_{p1} = \frac{i_{pp}}{\alpha}$$

where $i_{pp} = i_{p1} = -i_{p2}$ is the equivalent variational current that, if acting in the entire primary, would produce the same effects as the actual combination of i_{p1} and i_{p2} , and α is the ratio of the total secondary turns to total primary turns ($\alpha = N_2/N_1$).

The plate-to-plate load resistance, from a to b , Fig. 24.1, into

which the two tubes operate is given by

$$R_{pp} = \frac{R_L}{\alpha^2} \quad (24.7)$$

assuming the transformer to be ideal. With linear push-pull operation into an ideal output transformer with a pure resistance

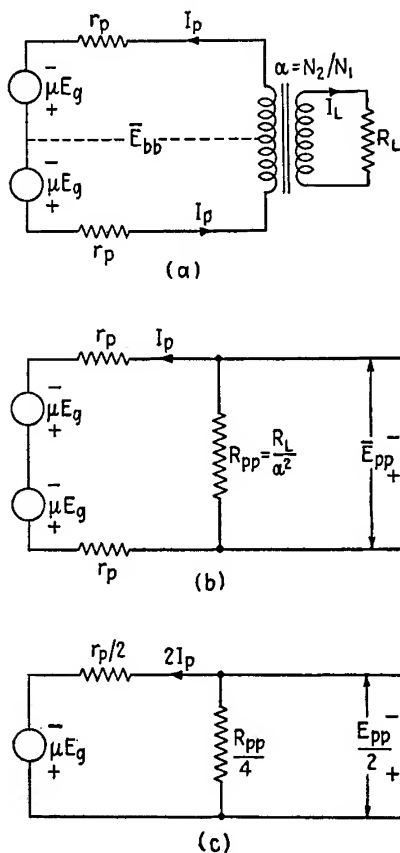


FIG. 24.2.—Equivalent plate circuits for push-pull linear operation.

R_L across its secondary, the load line for each tube must pass through its Q point and have a slope proportional to $-2/R_{pp}$.

The equivalent plate circuit of Fig. 24.1 for linear operation is given in Fig. 24.2a, where E_g is the value of grid-signal voltage to each tube. The variational plate resistance r_p and the amplification factor μ are measured at the Q point of each tube. The branch of the circuit containing \bar{E}_{bb} is shown dotted, Fig. 24.2a,

since no varying component of current flows in this branch, as was proved in (24.4). By replacing the ideal output transformer and load resistor with R_{pp} , the simplified equivalent plate circuit, Fig. 24.2b, is obtained, and the varying component of plate current is

$$I_p = \frac{2\mu E_g}{2r_p + R_{pp}} \quad (24.8)$$

The power delivered to R_{pp} and to the load resistor is given by

$$\tilde{P}_L = I_p^2 R_{pp} \quad (24.9)$$

or

$$\tilde{P}_L = \frac{4\mu^2 E_g^2}{(2r_p + R_{pp})^2} R_{pp} \quad (24.10)$$

The treatment of a push-pull stage can be made analogous to that of a single-sided stage incorporating a "composite tube" by reducing the equivalent circuit of Fig. 24.2b to that of Fig. 24.2c.

That the power delivered to the load impedance of Fig. 24.2c is the same as in Fig. 24.2b is verified by dividing both the numerator and the denominator of (24.10) by 16, which yields

$$\tilde{P}_L = \frac{\mu^2 E_g^2}{\left(\frac{r_p}{2} + \frac{R_{pp}}{4}\right)^2} \frac{R_{pp}}{4} \quad (24.11)$$

which is the equation for the power delivered to the resistor $R_{pp}/4$, Fig. 24.2c, from a generator having the emf μE_g and an internal impedance $r_p/2$. It should be noted that in Fig. 24.2c the load resistor and the voltage developed across it are the same as would be obtained if only one-half of the primary winding of the output transformer were used and that the varying component of plate current and the internal impedance of the generator are equivalent to placing the generators of Fig. 24.2b in parallel. Thus, this circuit of Fig. 24.2c is equivalent to the original push-pull stage only in that the two deliver the same power to the load resistor. The advantage of the composite-tube concept is more readily appreciated when the individual tubes are not restricted to linear operation.

Maximum power output, and (since no harmonic distortion is produced with linear operation) maximum undistorted power output, is obtained when R_{pp} is equal to $2r_p$.

The advantages of push-pull operation are not fully realized

when the operation of the tubes is restricted to the linear region. It is more important, therefore, to consider nonlinear Class A_1 operation in which the grid-signal voltage causes the plate current to be reduced below the upper limit of the region of large curvature, Fig. 23.1. Such conditions of operation are illustrated in Fig. 24.3 where the dynamic i_b - e_c characteristics of two identical tubes are shown with the characteristic of T_2 inverted and oriented so that the quiescent grid voltages \bar{E}_{c0} of the tubes coincide. The reason

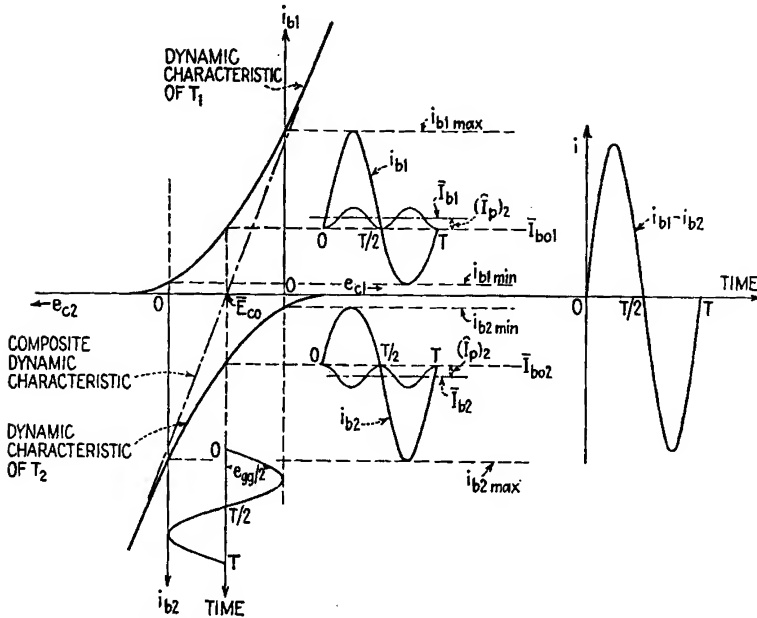


FIG. 24.3.—Graphical representation of Class A_1 nonlinear push-pull operation of two triodes.

for inverting the characteristic curve of T_2 is that the net effect in the primary of the output transformer is proportional to the difference between i_{b1} and i_{b2} , as in (24.5), and that, with the characteristic of T_2 inverted, i_{b2} appears on the graph as a negative quantity (*i.e.*, below the horizontal axis). The plate current for T_2 is plotted downward, and negative values of e_c are plotted to the right of the point denoting $e_{c2} = 0$. The varying component of grid voltage is the voltage applied to the grid circuit of either tube and is equal in magnitude to $e_{gg}/2$. The varying component of plate current for each tube is not sinusoidal, having, for the case shown, a fundamental and a second harmonic component. Since

the variational voltages applied to the grids have equal magnitudes but opposite signs,

$$e_{g2} = -e_{g1} \quad (24.12)$$

$$E_{g1} = \frac{E_{gg}}{2} \quad \text{and} \quad E_{g2} = -\frac{E_{gg}}{2} \quad (24.13)$$

Then, neglecting terms of higher order than the second in (3.1), the total instantaneous value of the plate current for T_1 is, with $e_{g1} = (E_{gg}/2) \sin \omega t$,

$$\begin{aligned} i_{b1} &= \bar{I}_{b0} + a \frac{\hat{E}_{gg}}{2} \sin \omega t + b \left(\frac{\hat{E}_{gg}}{2} \right)^2 \sin^2 \omega t \\ &= \bar{I}_{b0} + \frac{b}{2} \left(\frac{\hat{E}_{gg}}{2} \right)^2 + \frac{a\hat{E}_{gg}}{2} \sin \omega t - \frac{b}{2} \left(\frac{\hat{E}_{gg}}{2} \right)^2 \cos 2\omega t \end{aligned} \quad (24.14)$$

$$= \bar{I}_{b0} + (\bar{I}_p)_0 + (\hat{I}_p)_1 \sin \omega t - (\hat{I}_p)_2 \cos 2\omega t \quad (24.15)$$

where $(\bar{I}_p)_0$ is the rectified d-c component and $(\hat{I}_p)_1$ and $(\hat{I}_p)_2$ are the amplitudes of the fundamental and second harmonic components. Similarly, with $e_{g2} = -(\hat{E}_{gg}/2) \sin \omega t$,

$$i_{b2} = \bar{I}_{b0} + (\bar{I}_p)_0 - (\hat{I}_p)_1 \sin \omega t - (\hat{I}_p)_2 \cos 2\omega t \quad (24.16)$$

Substituting (24.15) and (24.16) into (24.6),

$$i_L = \frac{1}{\alpha} (\hat{I}_p)_1 \sin \omega t \quad (24.17)$$

and the second harmonic distortion does not appear in the load. From (24.4) the total instantaneous current through the power supply is

$$i_{bb} = 2\bar{I}_{bb} + 2(\bar{I}_p)_0 - 2(\hat{I}_p)_2 \cos 2\omega t \quad (24.18)$$

The average value of i_{bb} increases by $2(\bar{I}_p)_0$, which is equal to $2(\hat{I}_p)_2$, as in (24.14), (24.15). The varying component of current through the power supply consists of the second harmonic and all other even harmonics introduced by the tube characteristics. The fundamental frequency and all odd harmonics generated in the output stage are present in the output load current. The cancellation of the second harmonic distortion is made evident in Fig. 24.3, where the component sinusoids are shown in their actual phase relations. When the grid-signal voltage is nonsinusoidal, however, all frequencies included in it are present in the load current.

The equivalent varying component of current in the primary of the output transformer is the difference of the magnitudes of the two plate currents at each instant. This current may be

obtained also by projecting $e_{gg}/2$ on the composite i_b-e_c dynamic characteristic of the two tubes in push-pull.

The composite dynamic characteristic is obtained by taking the difference between the two plate currents for corresponding increments of grid voltages as measured from \bar{E}_{c0} , a positive increment for tube T_1 corresponding to an equal negative increment for tube T_2 . When the composite characteristic is a straight line as in Fig. 24.3, distortion produces only even harmonic frequencies in the individual plate currents and no distortion appears in the output. If, on the other hand, the individual tube characteristics are such that the composite characteristic is nonlinear, odd harmonics will be present in the output, the third harmonic being the most important. The amplitude of the third harmonic (not shown in Fig. 24.3) can be determined by the method discussed in Sec. 23 and the application of (23.17).

While the composite i_b-e_c diagram is useful for explaining the principle of operation of the push-pull amplifier, it is not convenient for design purposes because a new characteristic must be drawn for each value of plate-load resistance. An alternate method¹ for obtaining a family of composite i_b-e_b characteristics is useful for many applications.

These composite characteristics are based upon a composite tube, which is equivalent to the two tubes in the push-pull circuit. This treatment can be justified by the following considerations:

From Fig. 24.1,

$$e_{b1} = \bar{E}_{bb} - \frac{e_{pp}}{2} \quad (24.19)$$

$$e_{b2} = \bar{E}_{bb} + \frac{e_{pp}}{2} \quad (24.20)$$

The voltage developed across the load resistor R_L is

$$e_L = \alpha e_{pp} \quad (24.21)$$

also,

$$e_L = i_L R_L \quad (24.22)$$

Combining (24.22) and (24.6),

$$e_L = \frac{R_L}{2\alpha} (i_{b1} - i_{b2}) \quad (24.23)$$

From (24.23) and (24.21), the varying component of potential

¹ B. J. THOMPSON, Graphical Determination of Performance of Push-pull Audio Amplifiers, *Proc. I.R.E.*, **21**, 591, 1933.

developed across each half of the primary of the output transformer is

$$\frac{e_{pp}}{2} = \frac{R_L}{4\alpha^2} (i_{b1} - i_{b2}) \quad (24.24)$$

From (24.7),

$$\frac{e_{pp}}{2} = \frac{R_{pp}}{4} (i_{b1} - i_{b2}) \quad (24.25)$$

and (24.19) and (24.20) become

$$e_{b1} = \bar{E}_{bb} - \frac{R_{pp}}{4} (i_{b1} - i_{b2}) \quad (24.26)$$

$$e_{b2} = \bar{E}_{bb} + \frac{R_{pp}}{4} (i_{b1} - i_{b2}) \quad (24.27)$$

These equations show that the total instantaneous values of the plate potentials of T_1 and T_2 deviate by equal amounts from the quiescent value at any instant, the decrease in e_{b1} being equal to the increase in e_{b2} , and vice versa. These changes in plate potential are produced by increments of equal magnitude in e_{c1} and e_{c2} .

In Fig. 24.4, static i_b - e_b characteristics for T_1 and T_2 are shown for three values of grid potential, \bar{E}_{c0} , $\bar{E}_{c0} + e_{gg}/2$, $\bar{E}_{c0} - e_{gg}/2$, with e_{gg} assumed momentarily fixed. The characteristics of T_2 are drawn by rotating the characteristics of T_1 180° about the point denoting \bar{E}_{bb} on the e_b axis. The Q points of the tubes are denoted by the points Q_1 and Q_2 , the d-c resistance of the primary winding of the output transformer being neglected.

The composite static characteristic corresponding to zero signal voltage is obtained by determining the locus of values of $(i_{b1} - i_{b2})$ as e_{b1} and e_{b2} are changed. For example, in Fig. 24.4, when $e_{b1} = \bar{E}_{bb} - \Delta e_b$ and e_{b2} has the corresponding value $\bar{E}_{bb} + \Delta e_b$, the value of $i_{b \text{ comp}} = i_{b1} - i_{b2}$, as represented by the length \overline{cb} , is given by

$$\overline{cb} = \overline{ca} - \overline{cd} \quad (24.28)$$

Thus the composite static characteristic when $e_{gg} = 0$ is the locus of the points b corresponding to various values of Δe_b . Since a change of e_{c1} to $\bar{E}_{c0} + e_{gg}/2$ is accompanied by a change of e_{c2} to $\bar{E}_{c0} - e_{gg}/2$, the composite characteristic for $e_{c1} = \bar{E}_{c0} + e_{gg}/2$ (or $e_{c2} = \bar{E}_{c0} - e_{gg}/2$) is the locus of the points b' , obtained as in (24.28) with all letters primed. The locus of b'' is the composite static characteristic for $e_{c1} = \bar{E}_{c0} - e_{gg}/2$ (or $e_{c2} = \bar{E}_{c0} + e_{gg}/2$). A

complete family of composite static i_b - e_b characteristics of two triodes in push-pull operation is shown in Fig. 24.5.

The composite characteristics are much more nearly linear than the characteristics of either of the tubes but are not straight in all cases. They are, in many cases, parallel and equidistant for equal increments of grid-signal voltages.

The composite path of operation may be determined from (24.26) and (24.27), both of which represent a straight line passing through the point ($e_b = \bar{E}_{bb}$, $(i_{b1} - i_{b2}) = 0$) and having a negative slope determined by

$$\frac{i_{b1} - i_{b2}}{e_{b1} - \bar{E}_{bb}} = \frac{(i_b)_{\text{comp}}}{e_{pp}/2}$$

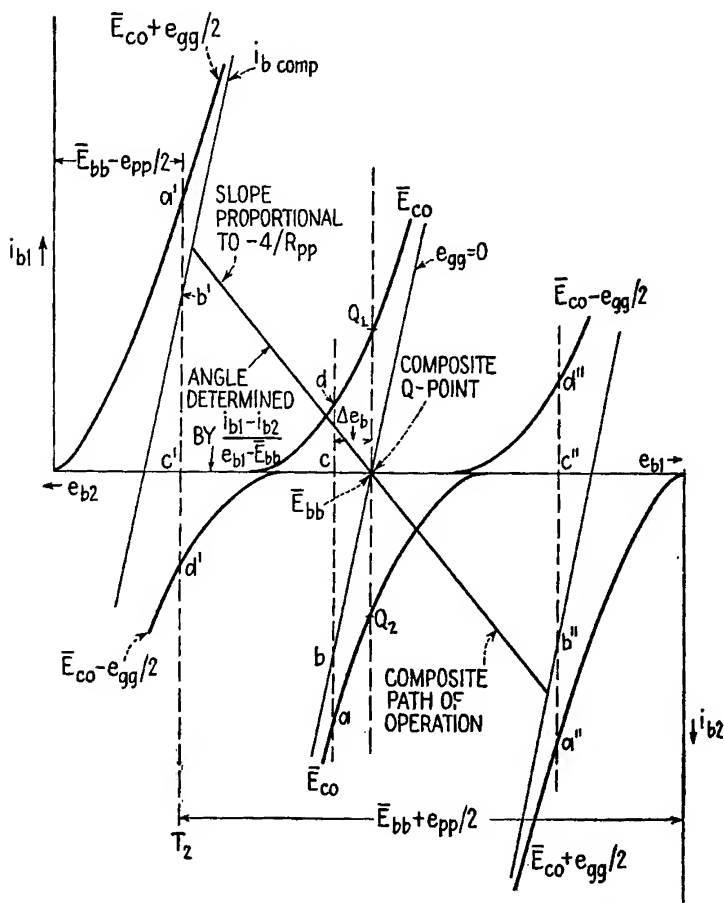


FIG. 24.4.—Determination of composite static characteristics.

where $i_{b \text{ comp}}$ is the total instantaneous value of the equivalent current in one-half the primary of the output transformer. Thus from (24.25) the composite path of operation has a slope proportional to

$$-\frac{4}{R_{pp}} = -\frac{4\alpha^2}{R_L}$$

as illustrated in Figs. 24.4 and 24.5. The plate-to-plate load resistance is, therefore, four times that represented graphically by the composite load line.

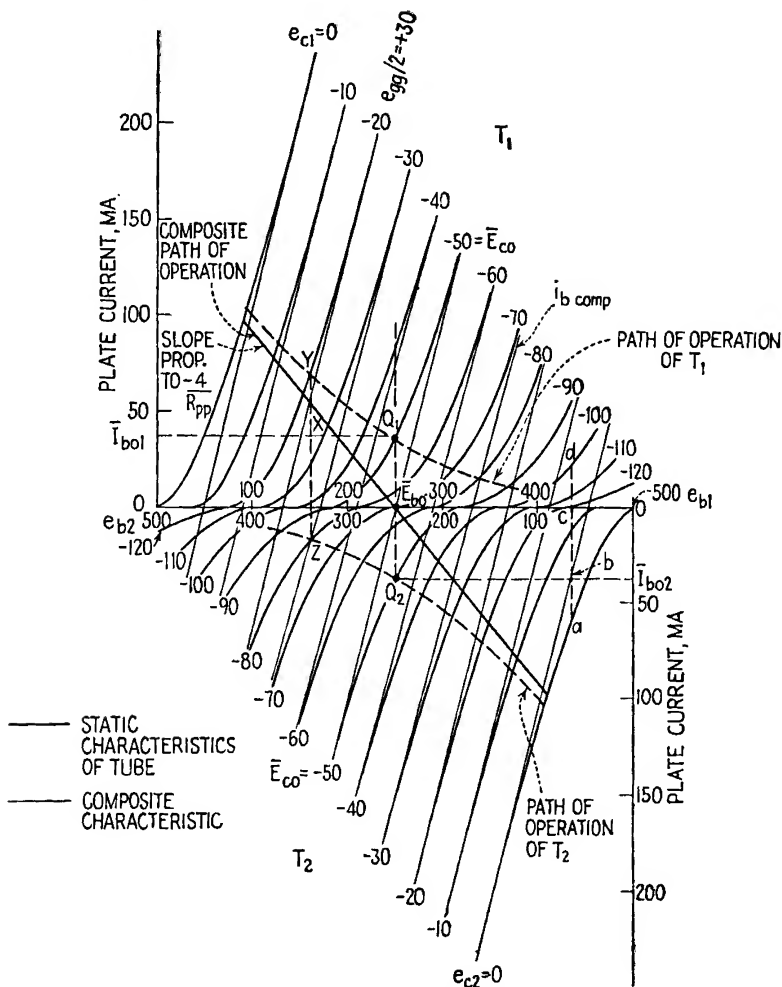


FIG. 24.5.—Composite $i_b - e_b$ characteristics of two triodes in Class A₁ nonlinear push-pull operation.

These considerations lead to an equivalent circuit for push-pull nonlinear operation similar to that of Fig. 24.2c for linear operation. This is possible because the composite i_b - e_c characteristics of the two tubes are approximately linear, equidistant, and parallel even though the individual tubes are operated beyond the region of linearity. Thus, let $r_{p \text{ comp}}$ be defined as the resistance corresponding to the reciprocal of the slope of the i_b $_{\text{comp}}$ characteristics and μ_{comp} as the negative of the ratio of Δe_b to Δe_c , i_b $_{\text{comp}}$ being constant. Then the circuit of Fig. 24.6 is applicable.

For maximum power output

$$\frac{R_{pp}}{4} = r_{p \text{ comp}}$$

i.e., the negative slope of the composite path of operation equals the slope of the composite static characteristics of the two tubes.

When the composite static characteristics are nonlinear, the equivalent circuit does not apply and one must employ graphical methods. In such cases, power output is often sacrificed, by using values of R_{pp} greater than $4r_{p \text{ comp}}$, in order that harmonic distortion may be decreased.

The path of operation for each tube is shown in Fig. 24.5 by heavy dashed lines. A vertical line drawn through the intersection X of the $e_{gg}/2 = +30$ -volt composite characteristics and the composite path of operation intersects the -20 -volt characteristic of T_1 at Y and the -80 -volt characteristic of T_2 at Z . The loci of points Y and Z are the paths of operation of the individual tubes. The curvature of these paths of operation with a resistive load is caused by operating the tubes over the nonlinear regions of their characteristics and by the interaction between the halves of the primary winding. The nonlinear action causes the instantaneous power output i_{pe_p} to be unequal for the two tubes.

Class A_1 operation exists so long as $i_{b \text{ min}}$ of either tube does not become zero before $i_{b \text{ max}}$ of the other tube is reached. Thus, the limiting condition of Class A_1 operation is that the grid-bias and grid-signal voltages be adjusted so that the plate current of one tube becomes zero at the time the total instantaneous grid voltage of the other tube reaches zero.

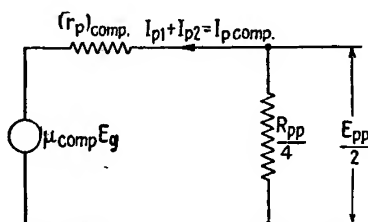


FIG. 24.6.—Equivalent plate circuit for a push-pull stage with nonlinear operation into an ideal output transformer.

The power output and plate-circuit efficiency of each tube and of the push-pull stage can be increased by adjusting the grid-bias voltage and the amplitude of the grid-signal voltage so that the

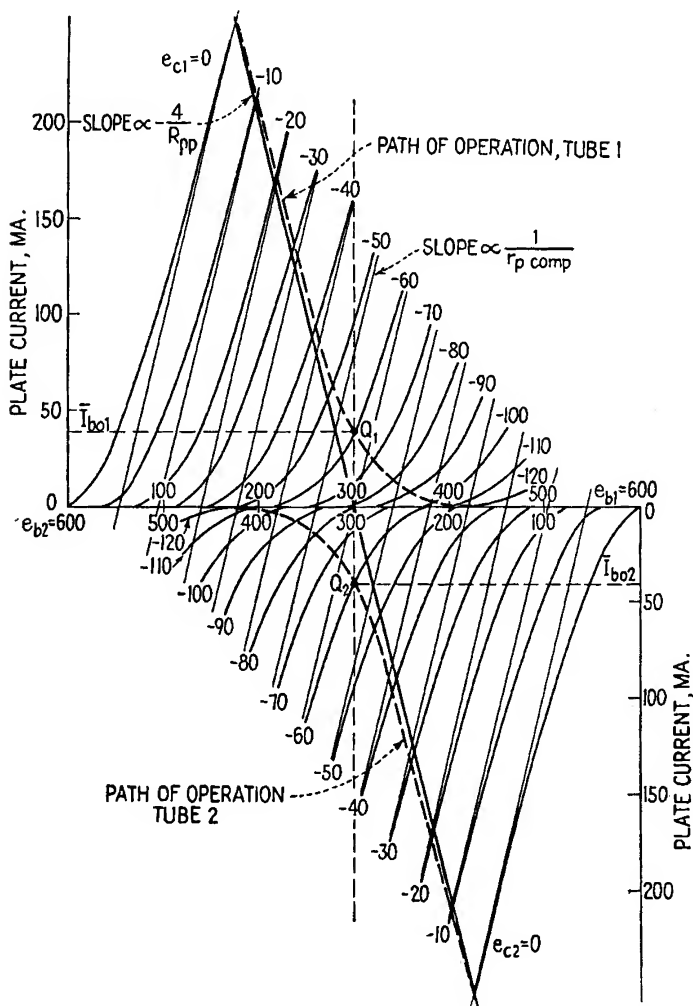


FIG. 24.7.—Composite i_b - e_b characteristics of two triodes in Class AB₁ push-pull operation.

plate current of each tube is zero for an appreciable part of each cycle. Class AB operation exists when plate current flows in each tube for appreciably more than one-half of the cycle, but not for the complete cycle. The composite i_b - e_b characteristics for Class

AB_1 operation of the tubes used in Fig. 24.5 are shown in Fig. 24.7. The plate-polarizing potential and grid-bias potential in this illustration are 300 volts and -60 volts. Even though the non-linear distortion produced by each tube is large, the equivalent primary current is very nearly sinusoidal. This is possible because the composite characteristics are essentially straight, parallel, and equally spaced for equal increments of grid voltage. The slope of the composite characteristics for Class AB operation is less than for Class A , equal values of \bar{E}_b being assumed, and the optimum load resistance for Class AB operation is greater than for Class A .

The average value of plate current for a tube operated Class AB differs from the quiescent value by the rectified component. The resultant increase in the d-c component of the plate current causes the power delivered by the power supply and, when cathode biasing is used, the negative grid-bias voltage to increase with the grid-signal voltage, necessitating good regulation of the plate power supply and a cathode by-pass capacitor if cathode degeneration is not desired.

Class AB_2 operation gives increased power output without an increase in power-supply voltage. Unless the impedance of the grid circuit is small, grid-circuit distortion occurs.

The power output and plate-circuit efficiency can be increased further by Class B operation. A balanced push-pull circuit cancels the even-harmonic distortion produced by each tube so that Class B operation is satisfactory for many a-f applications. By proper choice of plate-load resistance and grid-bias voltage the composite characteristics can be made approximately straight, parallel, and equidistant for equal increments of grid voltage.

In practice, the grid-bias voltage is chosen so that the linear portions of the dynamic i_b-e_c characteristics of each of the two tubes in push-pull fall on a common line, Fig. 24.8. The composite curve is usually slightly nonlinear in the region where both tubes are conducting. As \hat{E}_{gg} is increased, this region of nonlinearity becomes smaller in comparison with the range of operation, so that the distortion is unimportant for large values of \hat{E}_{gg} . For the purpose of this discussion, therefore, it is assumed that the composite i_b-e_c characteristic is linear as shown in the illustration.

The total instantaneous values of plate current for each tube of a Class B_2 push-pull stage are also plotted in Fig. 24.8. The current in the secondary of the output transformer is proportional to the algebraic sum of the variational components of individual

tube currents and is essentially sinusoidal, for a sinusoidal input voltage, if the composite i_b - e_c characteristic is linear.

The peaks of the plate-current curves are flattened if the equivalent grid-circuit impedance is appreciable. Hence the driving

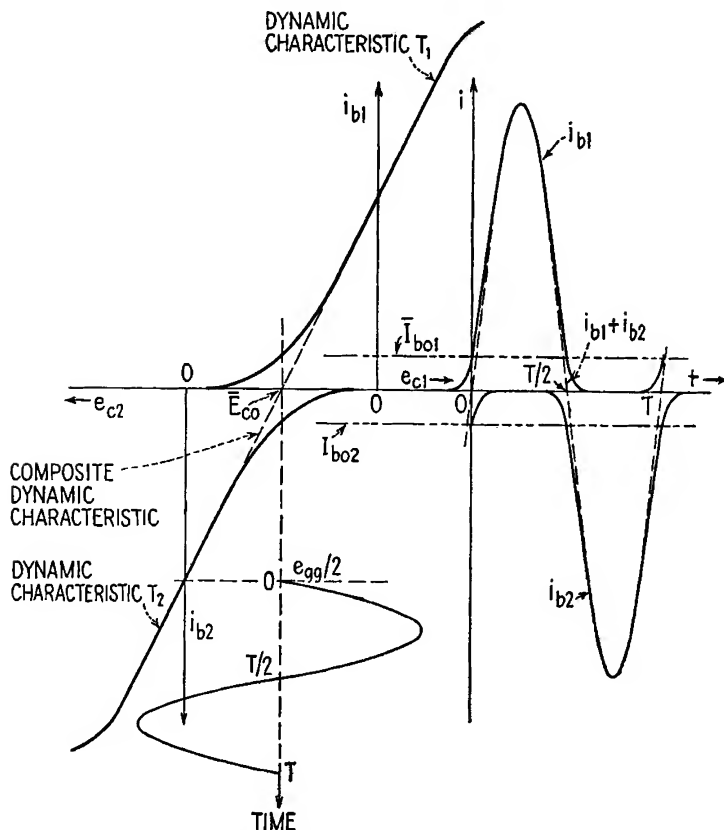


FIG. 24.8.—Composite i_b - e_c dynamic characteristics of two tubes in Class B_2 push-pull operation.

voltage to a Class B_2 stage is usually supplied from the driver stage by a step-down transformer. It is essential, also, that the driver stage be capable of supplying the power requirements of the grid circuits.

The large variations in the sum $i_{b1} + i_{b2}$ of the individual plate currents and the variation of the average values of the plate currents with signal voltage require good voltage regulation in the power supply. For the same reason, cathode biasing is not suited to

Class B operation. Some source of fixed bias voltage is required except in the case of specially designed tubes having a large μ and small plate current for which the grid-bias voltage is zero.

Since the grid driving power for Class AB_2 operation is usually smaller than for Class B_2 operation and because of the simplified problem of grid biasing, Class AB_2 output stages using beam-power tubes, which operate at slightly lower efficiency and much less nonlinear distortion, are generally preferred for audio systems except in high-powered installations.

25. Phase Inverters.—In Sec. 24, the input signal to the push-pull stage was supplied by an input transformer with a center-tapped secondary. Such transformers have a limited fre-

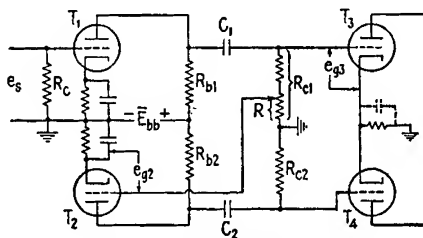


FIG. 25.1.—Two-tube phase inverter.

quency range, are heavy, occupy considerable space, and are relatively expensive. It is desirable in some applications, therefore, to drive a push-pull stage with resistance coupling when one side of the input signal is grounded. A number of circuits are available for this purpose; one is shown in Fig. 25.1.

This two-tube (T_1, T_2) phase-inverter circuit is essentially a two-stage resistance-coupled amplifier in which the grid-signal voltage to T_2 is obtained from the output of T_1 . When a positive increment of voltage is applied to the grid of T_1 , the potential of the grid of T_3 is made more negative with respect to its cathode. The signal voltage to the grid of T_2 is in phase with e_{g3} and is given by

$$e_{g2} = \frac{R}{R_{c1}} e_{g3}, \quad (25.1)$$

Consequently, a negative increment of voltage is applied to the grid of T_2 , causing the grid of T_4 to be made positive with respect to its quiescent voltage. If the value of R is adjusted so that R_{c1}/R is equal to the magnitude of the voltage amplification of T_2 and its circuit, grid-signal voltages equal in magnitude and 180° out

of phase are applied to the grids of T_3 and T_4 . The 180° -phase relation exists only in the mid-frequency range of the inverter stage (containing T_2). If the phase-inverter circuit is to be used for frequencies outside the mid-frequency range of the stage associated with T_1 , it is desirable to increase the mid-frequency range of the inverter stage to extend over the entire range of frequencies to be amplified. This can be accomplished by decreasing R_{b2} . The resulting decrease in voltage amplification in the inverter stage is readily overcome by increasing R to give an increased grid-signal voltage to T_2 .

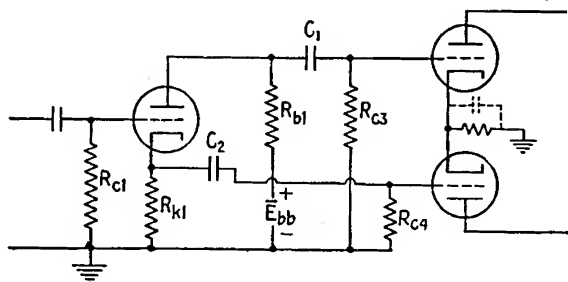


FIG. 25.2.—Single-tube phase-inverter circuit.

The grid-signal voltage to T_2 can be taken from a tap on R_{b1} if a blocking capacitor and grid resistor are inserted in the grid circuit of T_2 . A twin triode or a twin pentode is often used for T_1 and T_2 . A common cathode-biasing circuit for the tubes T_1 and T_2 often replaces the separate ones shown in the diagram.

The single-tube phase inverter employing cathode degeneration, Fig. 25.2, in which the inputs to the grids of the push-pull stage are taken from the plate and cathode, has the disadvantage of having a voltage amplification less than unity and of having unequal capacitances across the cathode and plate-load resistance.

26. Tuned Radio-frequency Amplifiers.—A modulated radio-frequency wave has a frequency spectrum which extends over a band of frequencies that is usually small when compared with the magnitude of the carrier frequency. It is necessary, therefore, to design r-f amplifiers which have large voltage amplification for frequencies in the neighborhood of the carrier frequency and which can be adjusted easily, within given limits, for various carrier frequencies. It is equally essential that r-f amplifiers have low voltage amplification for all frequencies outside the given band in order that a given station may be selected to the exclusion of all other stations which might be transmitting at that time. This

selectivity is accomplished by use of tuned parallel circuits in which the tuning capacitance is composed of the interstage wiring capacitance, the input and output capacitances of the tubes used, and whatever added capacitance may be required. By incorporating C_s in the tuned circuit, the voltage amplification at the resonant frequency of the plate-load impedance may be large even at high frequencies. The size of C_s may limit the maximum frequency to which the amplifier is tunable by means of the tuning capacitor.

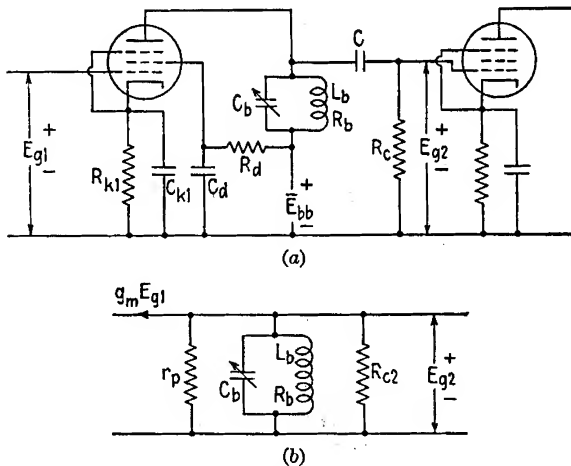


FIG. 26.1.—(a) Capacitance-coupled tuned-radio-frequency amplifier stage; (b) equivalent plate circuit.

The effect of the grid-to-plate capacitance C_{gp} is so large when triodes are used that neutralization usually is required. On the other hand, pentodes are almost universally used in radio-frequency voltage amplifiers thereby decreasing the effects of C_{gp} as well as the other tube capacitances and making larger voltage amplification possible. Three basic types of selective amplifiers, the capacitance-coupled, the tuned-secondary transformer-coupled, and the tuned-primary tuned-secondary transformer-coupled, are discussed in this order.

In the capacitance-coupled tuned-r-f amplifier, Fig. 26.1a, the plate-load impedance consists of a parallel combination of an inductor having inductance L_b and a-c resistance R_b and a variable capacitor having capacitance C_b . The coupling capacitance C has negligible reactance at radio frequencies, and the constant-current form of the equivalent plate circuit reduces to that in Fig. 26.1b. The output capacitance of the first tube, the input capaci-

tance to the second tube, and the interstage-wiring capacitance are incorporated in C_b . These capacitances plus the distributed capacitance of the inductor may be sufficient to require no added capacitance at high frequencies. If so, the tuning can be accomplished by L_b , usually made variable by use of an adjustable powdered-iron core.

When the carrier frequency is equal to the resonant frequency of the parallel circuit, the equivalent resistance R_t of the circuit is

$$R_t = \frac{L_b}{C_b R_b} = \omega_r L_b Q_r \quad (26.1)$$

where Q_r is the quality factor of the plate-tank circuit at the series-resonant frequency of $L_b C_b R_b$. It is assumed here that Q_r is large and that the parallel-resonant frequency does not differ appreciably from the series-resonant frequency.¹

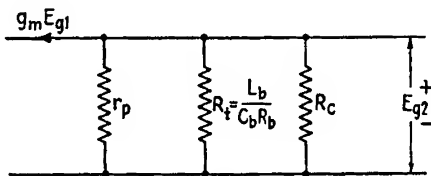


FIG. 26.2.—Equivalent plate circuit of Fig. 26.1a when the plate circuit is tuned to parallel resonance.

When Q_r is large and the plate-tank circuit is tuned to parallel resonance, therefore, the circuit of Fig. 26.1b reduces to that in Fig. 26.2. Since the amplification $A_r = -g_m/Y_L$ and

$$Y_L = \frac{1}{r_p} + \frac{1}{R_t} + \frac{1}{R_c}$$

at parallel resonance, the maximum voltage amplification is given by

$$A_r = \frac{-g_m}{\frac{1}{r_p} + \frac{1}{R_t} + \frac{1}{R_c}} \quad (26.2)$$

which combined with (26.1) gives

$$A_r = \frac{-g_m \omega_r L_b Q_r}{1 + \frac{R_t}{r_p} + \frac{R_t}{R_c}} \quad (26.3)$$

¹ The expression for the impedance of the tank circuit when Q_r is small is discussed in Chap. II, Sec. 15. This condition is not often encountered in r-f amplifiers.

If r_p and R_c are both large compared with R_t ,

$$A_r = -g_m \omega_r L_b Q_r \quad (26.4)$$

which suggests the possibility of rewriting (26.3) as

$$A_r = -g_m \omega_r L_b Q_r' \quad (26.5)$$

in which the "loaded" Q of the plate circuit

$$Q_r' = \frac{Q_r}{1 + \frac{R_t}{r_p} + \frac{R_t}{R_c}} \quad (26.6)$$

is the quality factor of the complete plate circuit, the effects of r_p and R_c being taken into consideration.

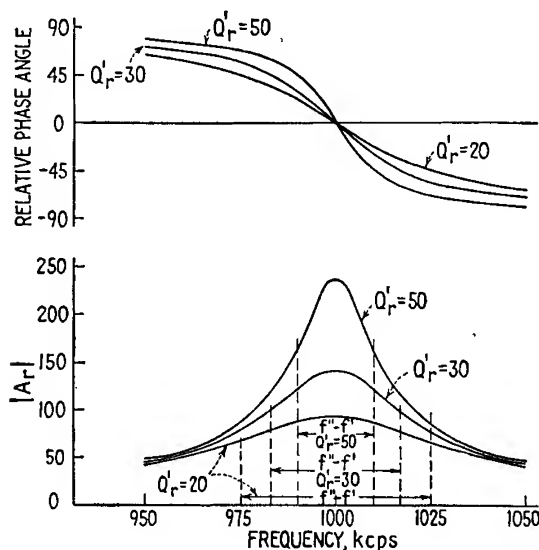


FIG. 26.3.—Frequency and phase response near resonance for a capacitance-coupled r-f amplifier stage.

The band width of one stage is

$$f'' - f' = \frac{f_r}{Q_r'} \quad (26.7)$$

The effect of r_p and R_c is to reduce the effective Q of the plate circuit, thus increasing the band width of the stage as R_c (or r_p) is decreased and decreasing the voltage amplification.

The circuit of Fig. 26.1a may be changed so that the tank circuit

replaces R_c and a plate-load resistor replaces the parallel circuit without altering the treatment just given.

The curves of Fig. 26.3 show the variations in the voltage amplification, band width, and phase angle, as a function of loading.

27. Tuned-secondary Transformer-coupled Amplifiers.—The typical circuit diagram of a tuned-secondary transformer-coupled amplifier stage is given in Fig. 27.1a. This type of amplifier stage is known also as a *tuned-radio-frequency* (TRF) stage. The

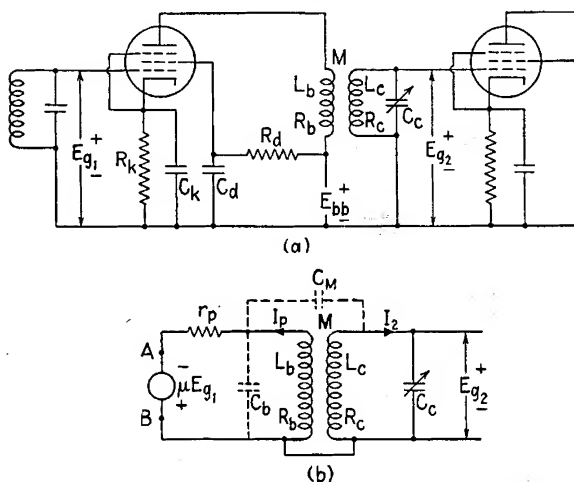


FIG. 27.1.—(a) Tuned-secondary transformer-coupled amplifier stage; (b) equivalent circuit.

equivalent plate-circuit diagram employing a constant-voltage generator is shown in Fig. 27.1b. The output capacitance of the tube and the distributed capacitance of the primary are combined in C_b , which is in shunt with the primary. The distributed capacitances between primary and secondary of the transformer are lumped in the capacitance C_M . At broadcast frequencies, C_b and C_M may be neglected ordinarily. C_b and C_M being neglected, the total impedance into which the generator is working is

$$Z_{AB} = Z_{11} + \frac{\omega^2 M^2}{Z_{22}} \quad (27.1)$$

where Z_{11} is the total series impedance of the primary circuit, Z_{22} is the total impedance of the secondary circuit, M the mutual inductance between primary and secondary, and ω the angular frequency of the signal voltage. Therefore,

$$Z_{11} = r_p + R_b + j\omega L_b$$

and

$$Z_{22} = R_c + j \left(\omega L_c - \frac{1}{\omega C_c} \right) \quad (27.2)$$

Substituting from (27.2) in (27.1),

$$Z_{AB} = r_p + R_b + j\omega L_b + \frac{\omega^2 M^2}{R_c + j \left(\omega L_c - \frac{1}{\omega C_c} \right)} \quad (27.3)$$

The alternating component of the plate current is

$$I_p = \frac{\mu E_{g1}}{Z_{AB}} \quad (27.4)$$

and the induced emf in the secondary is

$$E_2 = -j\omega M I_p \quad (27.5)$$

The secondary current is

$$I_2 = \frac{E_2}{Z_{22}} \quad (27.6)$$

and the signal voltage to the grid of the second tube is

$$E_{g2} = -j \frac{I_2}{\omega C_c} \quad (27.7)$$

Combining equations (27.3) to (27.7),

$$A = \frac{E_{g2}}{E_{g1}} = \frac{-\omega M \mu / \omega C_c}{\left[R_c + j \left(\omega L_c - \frac{1}{\omega C_c} \right) \right] \left[r_p + R_b + j\omega L_b + \frac{\omega^2 M^2}{R_c + j \left(\omega L_c - \frac{1}{\omega C_c} \right)} \right]} \quad (27.8)$$

Since this is a selective amplifier, the voltage amplification at or near the resonant frequency of the secondary is of primary interest. At the resonant frequency of the secondary,

$$\omega_r^2 = \frac{1}{L_c C_c} \quad (27.9)$$

and (27.8) reduces to

$$A_r = - \frac{\omega_r M \mu}{\omega_r C_c R_c \left(r_p + R_b + j\omega_r L_b + \frac{M^2 \omega_r^2}{R_c} \right)} \quad (27.10)$$

The impedance of the primary of the transformer usually is

small compared with the variational plate resistance of the tube. Equation (27.10) reduces, therefore, to

$$A_r \doteq - \frac{\omega_r M \mu}{\omega_r C_c R_c \left(r_p + \frac{M^2 \omega_r^2}{R_c} \right)} \quad (27.11)$$

when $r_p > (R_b + j\omega_r L_b)$.

Dividing both numerator and denominator of (27.11) by r_p and substituting $Q_c = \omega_r L_c / R_c = 1 / \omega_r C_c R_c$,

$$A_r = -g_m \frac{\omega_r M Q_c}{1 + \frac{\omega_r^2 M^2}{r_p R_c}} \quad (27.12)$$

In some applications, when the tube is a pentode, r_p is of the order of 1 megohm and $\omega_r^2 M^2 / r_p R_c \ll 1$. In such cases,

$$A_r \doteq -g_m \omega_r M Q_c \quad (27.13)$$

The conditions for maximum voltage amplification at resonance in a TRF stage may be obtained by equating to zero the derivative of A_r in (27.12) with respect to M . This gives

$$\omega_r M_1 = \sqrt{r_p R_c} \quad (27.14)$$

where M_1 is the value of M for which the resistance reflected from the secondary circuit into the primary circuit is equal to the variational plate resistance of the tube. It is the condition for maximum power transfer if linear operation of the tube is assumed. The maximum voltage amplification is, therefore,

$$|A_{r \max}| \doteq \frac{g_m Q_c \sqrt{r_p R_c}}{2} \quad (27.15)$$

When pentodes are employed, it is impossible to obtain values of M large enough to satisfy the conditions for maximum voltage amplification in the broadcast band of frequencies.

The variation of the voltage amplification as a function of M is illustrated in Fig. 27.2a for three resonant frequencies (fr_1, fr_2, fr_3) near the lower, the middle, and the upper limit of the broadcast band. It will be noted that the value of M for maximum voltage amplification decreases as the frequency to which the stage is tuned increases and that voltage amplification increases with frequency. The curves of Fig. 27.2b illustrate the dependence of the voltage

amplification at frequencies near resonance upon the mutual inductance.

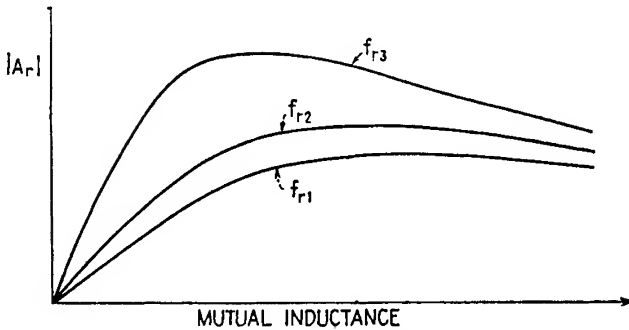


FIG. 27.2.—(a) Effect of mutual inductance on $|A_r|$ in a tuned-secondary transformer-coupled amplifier stage; $f_{r3} > f_{r2} > f_{r1}$.

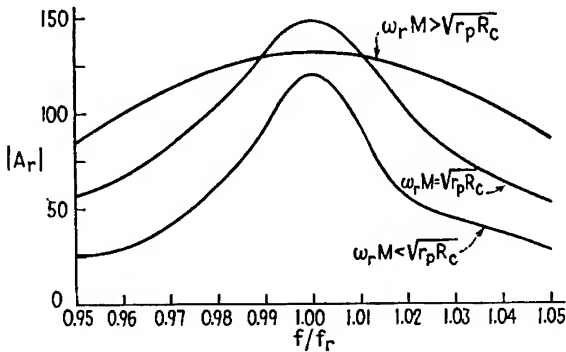


FIG. 27.2.—(b) Effect of mutual inductance on the frequency response of a tuned-secondary transformer-coupled amplifier stage.

28. Tuned-primary Tuned-secondary Transformer-coupled Amplifiers.—The tuned-primary tuned-secondary transformer-coupled amplifier stage differs from the TRF stage in that the primary as well as the secondary of the transformer is tuned to the carrier frequency of the signal. This type of amplifier is used almost exclusively in superheterodyne receivers for amplifying the intermediate-frequency signal. Since in any given receiver the intermediate frequency is fixed, provision for continuous tuning is not provided. Adjustable capacitors or inductors are available, however, for aligning the various stages.

From the typical circuit diagram and the equivalent plate circuit employing a constant-voltage generator, Fig. 28.1, it is evident that the primary capacitance C_b is the sum of the output

capacitance C_o of the first tube, the distributed capacitance of the primary, and any added capacitance; and the secondary capacitance C_e consists of the distributed capacitance of the secondary coil and wiring, the input capacitance of the second tube, and the added capacitance for tuning. The distributed capacitance between the primary and secondary is represented by the lumped capacitor C_M . The capacitance C_M is usually small and is neglected in this treatment.

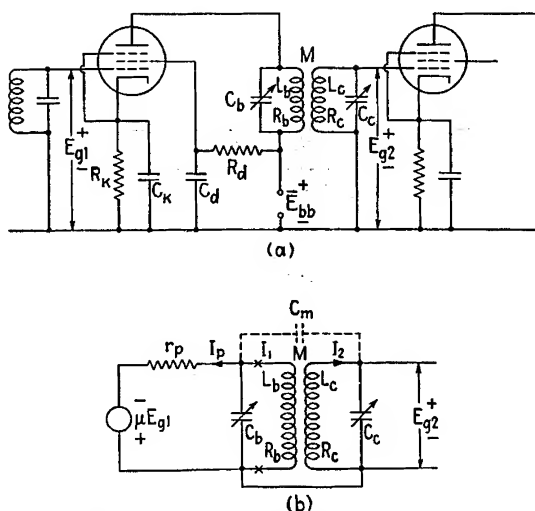


FIG. 28.1.—(a) Tuned-primary tuned-secondary transformer-coupled amplifier stage; (b) equivalent circuit.

By applying Thévenin's theorem¹ to that part of the circuit, Fig. 28.1b, to the left of xx , the primary circuit may be transformed to that of Fig. 28.2a. In many cases $\omega^2 C_b^2 r_p^2 \gg 1$, and the modified equivalent circuit of Fig. 28.2a may, for practical applications, be simplified to the approximate form of Fig. 28.2b. In this form the circuit is that of an ordinary two-mesh magnetically coupled circuit, Chap. VII. The alternating current I_1 is the current in the primary of the transformer rather than the alternating plate current of the tube. Applying the coupled-circuit analysis in the same manner as in Sec. 27,

$$Z_{11} = \frac{1}{\omega^2 C_b^2 r_p} + R_b + j \left(\omega L_b - \frac{1}{\omega C_b} \right) \quad (28.1)$$

$$Z_{22} = R_c + j \left(\omega L_c - \frac{1}{\omega C_c} \right) \quad (28.2)$$

¹ See Chap. V, Sec. 5.

and

$$Z_{AB} = Z_{11} + \frac{\omega^2 M^2}{Z_{22}} \quad (28.3)$$

It follows, therefore, that

$$I_1 = \frac{-j\mu E_{g1}}{\omega C_b r_p Z_{AB}} \quad (28.4)$$

$$I_2 = -\frac{j\omega M I_1}{Z_{22}} \quad (28.5)$$

and

$$E_{g2} = -j \frac{I_2}{\omega C_c} \quad (28.6)$$

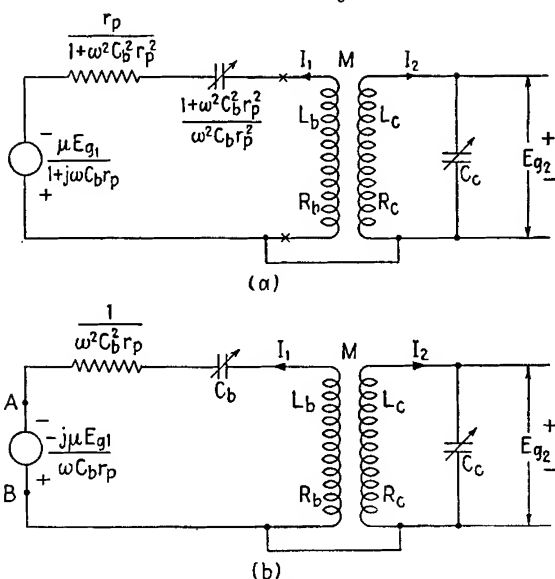


FIG. 28.2.—Simplified equivalent plate circuits for a tuned-primary tuned-secondary transformer-coupled amplifier stage.

Combining (28.1) to (28.6),

$$A_r = \frac{E_{g2}}{E_{g1}} = \frac{+jg_m \omega_r M}{\omega_r^2 C_c C_b R_c \left(\frac{1}{\omega_r^2 C_b^2 r_p} + R_b + \frac{\omega_r^2 M^2}{R_c} \right)} \quad (28.7)$$

when both primary and secondary circuits are tuned to resonance so that

$$\omega_r^2 = \frac{1}{L_c C_c} = \frac{1}{L_b C_b} \quad (28.8)$$

Equation (28.7) may be made more useful by expressing it in terms of the coefficient of coupling k , the quality factor Q_b' of the

primary, and the quality factor Q_c of the secondary, these being defined by

$$k = \frac{|M|}{\sqrt{L_b L_c}} \quad (28.9)$$

$$Q_b' = \frac{\omega_r L_b}{\frac{1}{\omega_r^2 C_b^2 r_p} + R_b} = \frac{1}{\omega_r C_b \left(\frac{1}{\omega_r^2 C_b^2 r_p} + R_b \right)} \quad (28.10)$$

and

$$Q_c = \frac{\omega_r L_c}{R_c} = \frac{1}{\omega_r C_c R_c} \quad (28.11)$$

The magnitude of the voltage amplification is

$$|A_r| = \frac{g_m k \omega_r \sqrt{L_b L_c}}{\frac{1}{Q_b' Q_c} + k^2} \quad (28.12)$$

The coefficient of coupling for maximum voltage amplification at resonance may be determined by equating to zero the derivative of (28.12) with respect to k , whence

$$k_c^2 = \frac{1}{Q_b' Q_c} \quad (28.13)$$

where k_c is the coefficient of *critical* coupling. When critical coupling is obtained, the conditions for maximum power transfer exist and the resistance reflected into the primary circuit from the secondary circuit is equal to the resistance of the primary circuit.

The frequency response of an i-f stage in the neighborhood of the resonant frequency is a function of the coefficient of coupling between the primary and secondary. When Q_b' equals Q_c , the maximum coupling for a single-peak response curve is critical coupling. The band width of the response curve for the condition of critical coupling is $\sqrt{2}$ times the band width of a single resonant circuit. The band width may be increased by increasing the coupling above the critical value; but the voltage amplification at and near the resonant frequency decreases, and the phase relations are not so good. Figure 28.3 shows measured voltage-response curves of an intermediate-frequency amplifier stage having approximately equal Q 's in the primary and secondary circuits. These curves should be compared with those of Fig. 16.1, Chap. VII. The shape of the voltage-response curves differs slightly from that of the current-response curves (Fig. 16.1, Chap. VII) since the

reactance of the secondary capacitance decreases as the frequency increases, causing the voltage peaks to be somewhat lower at the higher frequencies than at the lower frequencies. Also, it must be borne in mind that experimentally determined response curves usually differ from the calculated, or "theoretical," curves because the calculations ordinarily are based upon simplifying assumptions. In a radio receiver the tuned circuits may be more or less mis-

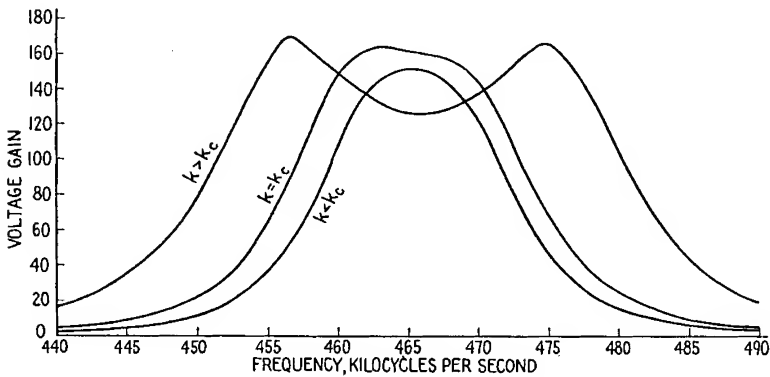


FIG. 28.3.—Measured frequency-response curves of an intermediate-frequency amplifier stage.

aligned (*i.e.*, not all tuned *exactly* to the desired frequency), and there are usually some coupling and regeneration due to interelectrode capacitances of tubes and distributed capacitances between circuit elements and circuit wiring.

In most applications Q_c is larger than Q_b' . Maximum voltage amplification is obtained when the coupling is critical irrespective of the values of Q_b' and Q_c , but it is possible to increase the coupling above critical without obtaining a double-peaked response curve when Q_b' is not equal to Q_c . The maximum coefficient of coupling for a single-peaked response curve when $Q_b' \neq Q_c$ is known as the *optimum coupling* and is given by

$$k_o = \sqrt{\frac{1}{2} \left(\frac{1}{Q_b'^2} + \frac{1}{Q_c^2} \right)} \quad (28.14)$$

When $Q_b' = Q_c$, the coefficient of critical coupling is identical with optimum coupling. The voltage amplification at the resonant frequency is less with optimum coupling than with critical coupling. The main advantage of optimum coupling is the increased band width.

The dependence of the frequency response on the coefficient of coupling when $Q_b' < Q_c$ is illustrated by the current-response curves of Fig. 16.2, Chap. VII, where Q_1 corresponds to Q_b' and Q_2 corresponds to Q_c . When $Q_b' < Q_c$ and the coefficient of coupling is greater than critical, maximum or peak voltage amplification is always less than that obtained at the resonant frequency with critical coupling. On the other hand, when $Q_b' = Q_c$ and the coupling is greater than critical, the maximum or peak voltage amplification occurs at two frequencies and is approximately equal to that obtained at the resonant frequency with critical coupling.

Tuned-primary tuned-secondary transformer-coupled amplifiers in which the coupling is critical or optimum have a frequency-response curve that is much flatter near the resonant frequency than is possible with single tuned circuits. The increased steepness of the sides of the response curve provides greater selectivity than is possible with single circuits.

29. Feedback in Amplifiers.—When a part of the input voltage of an amplifier stage is obtained from the output of that stage or following stages, feedback action is present. Feedback occurs either through a path that is inherent in the system or through a *feedback circuit* specially designed for the purpose. When feedback aids or opposes the amplification, it is called *positive* or *negative*, respectively.

Positive feedback in an amplifier may result in sustained oscillations. An amplifier in which there is positive feedback, but not sufficient feedback to cause sustained oscillations, is said to be *regenerative*.

Negative feedback may be employed in a low-frequency amplifier in order to achieve one or more of several desirable results described below. Such an amplifier is said to be *degenerative*. Negative feedback is used to eliminate the effects of the positive feedback inherent in a high-frequency amplifier due to the grid-to-plate capacitance. This use of negative feedback is known as *neutralization*, Chap. XIV.

The action of a feedback path depends upon the frequency of the variational voltage. An amplifier that has feedback may remain either regenerative or degenerative throughout the range of signal frequencies that are applied, though the magnitude and phase angle of the feedback voltage may change with frequency. An amplifier may be regenerative in one frequency range and degenerative in another.

Generally, positive feedback tends to sharpen the frequency-response curve and to decrease the range of uniform response. Therefore, regeneration, or positive feedback, may be employed to increase the gain and selectivity of an amplifier.¹ Negative feedback tends to flatten the frequency-response curve and extend the range of uniform response. It also reduces harmonic (non-linear) and phase distortion. Negative feedback increases an amplifier's stability, making the voltage amplification less dependent upon operating voltages and tube coefficients. Under certain conditions it also decreases the noise in the output of an amplifier. Also, feedback networks of special design may be employed to provide

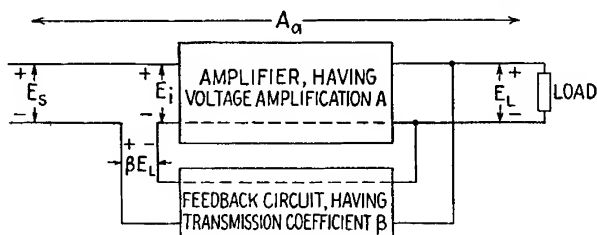


FIG. 29.1.—Principle of feedback in amplifiers.

selective attenuation or emphasis to compensate for undesired frequency characteristics in circuits external to the amplifier.²

Feedback may be harmful. A small amount of positive feedback in a band-pass amplifier may cause distortion. A large amount of positive feedback in any system may result in oscillation. Prevention of undesired feedback is a primary purpose of such arrangements as the resistor-capacitor "decoupling filter" used in plate and grid-supply circuits.³

Figure 29.1 illustrates the principle of feedback in an amplifying system. For simplicity, series injection is shown at the input, but other forms of network coupling may be employed. The feedback circuit has a transmission coefficient β , which is the ratio of the feedback voltage to the load voltage. The voltage amplification of the amplifier alone is defined as

$$A = \frac{E_L}{E_i} \quad (29.1)$$

while the amplification of the system with feedback, called the

¹ Chap. XXIII, Sec. 2.

² H. S. BLACK, *Stabilized Feedback Amplifiers*, *Trans. A.I.E.E.*, **53**, 114, January, 1934; U.S. patent 2,106,671, Dec. 21, 1937.

³ See Sec. 33.

actual voltage amplification, is

$$A_a = \frac{E_L}{E_s} \quad (29.2)$$

Each of the quantities A , A_a , β , in general, is complex and is a function of frequency. The input voltage E_i is equal to $E_s + \beta E_L$. Substitution in (29.2) of this value of E_i together with the value of E_L from (29.1) gives

$$A_a = \frac{A}{1 - \beta A} \quad (29.3)$$

This is the fundamental feedback formula. The product βA is defined as the *feedback factor*. By definition: with positive feedback,

$$|A_a| > |A| \quad \text{and} \quad |1 - \beta A| < 1 \quad (29.4)$$

with negative feedback,

$$|A_a| < |A| \quad \text{and} \quad |1 - \beta A| > 1 \quad (29.5)$$

Equation (29.3) is applicable when the system is acting as a linear amplifier under steady-state conditions.

When feedback is employed, there is always danger that oscillation will occur at some frequency. The conditions necessary for oscillation have been investigated.¹ Let the feedback factor βA and its conjugate be plotted on the complex plane as a function of frequency. If this locus encloses or includes the point $1 + j0$, the system oscillates and will not act as a stable amplifier.

As positive feedback is increased from zero, (29.3) applies until the amplitude of the output voltage becomes so large that non-linearity results. If the positive feedback is increased sufficiently, the amplifier usually oscillates. The amplitude of the oscillatory voltage is limited by tube and circuit conditions; when the rate of energy dissipation becomes equal to the rate of energy input, the oscillations become steady. That is, the steady oscillatory state is reached when the negative resistance developed in the amplifier is equal to the total positive resistance.

If the positive feedback is further increased, the oscillations usually continue, the waveform differing more and more from

¹ H. NYQUIST, Regeneration Theory, *Bell System Tech. J.* **11**, 126, January, 1932; E. PETERSON, F. G. KREER, and L. A. WARE, Regeneration Theory and Experiment, *Proc. I.R.E.*, **22**, 1191, October, 1934; H. W. BODE, Relations between Attenuation and Phase in Feedback Amplifier Design, *Bell System Tech. J.*, **19**, 421, July, 1940.

that of a sinusoid and the fundamental frequency differing more and more from the value at the inception of oscillation. In many cases, extreme positive feedback causes the amplifier to behave like a relaxation oscillator, Chap. XV.

With negative feedback and the feedback factor βA having a magnitude large compared with unity, (29.3) becomes

$$A_a \doteq -\frac{1}{\beta} \quad (29.6)$$

This important approximation shows that, when $|\beta A| \gg 1$, the actual amplification with negative feedback is a function of the characteristics of the feedback network only and is practically independent of such factors as the tube characteristics and the polarizing voltages. Equation (29.6) indicates that β must be small for A_a to be large. Then the value of A must be large in order to have $|\beta A| \gg 1$. When the conditions that underlie (29.6) are satisfied, the actual voltage amplification may be made independent of frequency by employing a resistive feedback network.

The effect of a large amount of negative feedback may be explained qualitatively as follows: E_s and βE_L are of the same order of magnitude. If, owing to variation of the voltage amplification A with frequency or for any other cause, the load voltage E_L tends to be changed in value, the feedback voltage βE_L is changed proportionally. Any variation in the latter voltage, however, causes considerable change in the input voltage E_i , since E_i is the small difference between E_s and βE_L . This change produces an effect at the output that opposes and practically cancels the original variation in load voltage E_L , which, therefore, remains essentially constant.

The frequency response of the system may be altered by making the feedback network reactive, so that β is smaller for those frequencies for which increased amplification is desired, and larger for those frequencies to be attenuated.

30. Current- and Voltage-controlled Feedback.—The variational voltage fed back from the output of an amplifier may be proportional either to the *voltage across* the load or to the *current through* the load, and accordingly the feedback may be defined as voltage-controlled or current-controlled, Fig. 30.1. In Fig. 30.1a the feedback voltage developed between points 4 and 5 results from voltage division of the voltage across the load, while in Fig. 30.1b it results from load current passing through a resistance

in the output circuit. In either of these examples, the amplifier can be either regenerative or degenerative, depending upon the connection of the output transformer. An amplifier may possess any combination of voltage- and current-controlled feedback.

Another form of current-controlled negative feedback, known as *cathode degeneration*, Sec. 31, occurs when there is an impedance

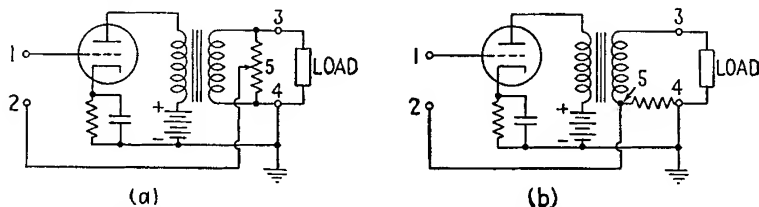


FIG. 30.1.—Examples of an amplifier with (a) voltage-controlled feedback and (b) current-controlled feedback.

in the cathode circuit. For example, if the by-pass capacitor is omitted or is insufficient when cathode bias is employed, cathode degeneration results.

In the circuit of Fig. 30.2, voltage feedback from the output is secured by the voltage divider formed by R_1 and R_2 . R_2 also causes cathode degeneration. The voltage developed across R_2 by

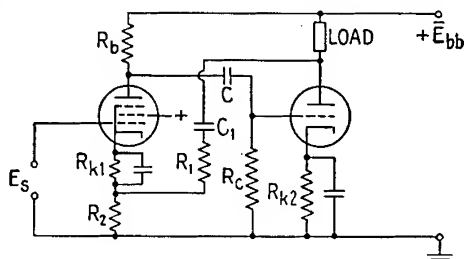


FIG. 30.2.—Two-tube amplifier with over-all voltage-controlled negative feedback and cathode degeneration.

division of the output voltage opposes the input-signal voltage, so that this feedback is negative.

Feedback may result in considerable alteration of the input and output impedances of an amplifier. For example, with voltage-controlled negative feedback, the output impedance is smaller than it would be without feedback. With current-controlled negative feedback the output impedance is increased.

The alteration of output impedance may be accounted for as follows: Figure 30.3a represents an amplifier with voltage-controlled negative feedback. The output stage of the amplifier is represented

by its equivalent plate circuit in which an emf kE_i is acting in series with the equivalent internal impedance Z , which would be the output impedance of the amplifier if no feedback were present. The factor k is a complex number representing the amplification from the input to the amplifier up to the input to the final stage. The voltage divider R_1R_2 provides the feedback voltage; for simplicity, let the resistance of this voltage divider be very much greater than the impedance Z_L of the load. The negative feedback voltage βE_L tends to decrease the magnitudes of E_i , kE_i , I_L , and E_L .

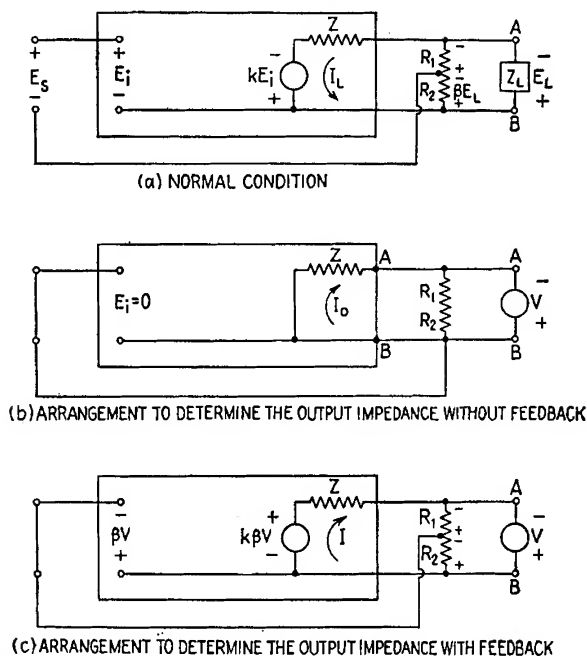


FIG. 30.3.—Output impedance of an amplifier with voltage-controlled negative feedback.

To determine the output impedance without feedback, the circuit of Fig. 30.3b is used. The input terminals are short-circuited; the tap on R_1R_2 is moved to the bottom of R_2 ; and an a-c source of voltage V , having the signs shown, is put in place of the load Z_L . (The current I_o that results is opposite in direction to that which would be produced by the voltage E_s in Fig. 30.3a.) The output impedance without feedback is

$$\frac{V}{I_o} = Z \quad (30.1)$$

To determine the output impedance with feedback, the feedback tap is replaced in its original position, and the conditions shown in Fig. 30.3c result. The current due to V is now

$$I = \frac{V + k\beta V}{Z} = \frac{V(1 + k\beta)}{Z} \quad (30.2)$$

and the output impedance with feedback is

$$Z_{AB} = \frac{V}{I} = \frac{Z}{1 + k\beta} \quad (30.3)$$

which indicates that with negative voltage-controlled feedback, $Z_{AB} < Z$, or the output impedance with feedback is less than the output impedance without feedback.

If the sign of either β or k is changed, the feedback becomes positive, and the output impedance is increased instead of decreased. The effects of positive and negative current-controlled feedback upon output impedance may be worked out in corresponding fashion.

With the output impedance of the amplifier decreased, the load voltage E_L is less dependent upon the value of the load impedance, hence less dependent upon frequency. The amplifier then may be compared with a constant-voltage generator. With the output impedance increased the current through the load is less dependent upon load impedance, so that the amplifier then may be compared with a constant-current generator.

31. Cathode Degeneration.—A simple circuit illustrating cathode degeneration is shown in Fig. 31.1a. Since the output voltage E_o is the total alternating potential developed across the plate-load impedance Z_L , it does not represent the voltage delivered to the following stage in the general case. When the d-c potential developed across Z_k differs from the desired grid-polarizing potential, methods similar to those shown in Fig. 32.3 may be used to obtain the proper value of grid-bias voltage.

Analysis of the circuit of Fig. 31.1a results in the equivalent plate-circuit diagram of Fig. 31.1b, from which

$$I_p = \frac{\mu(E_s - E_k)}{r_p + Z_L + Z_k} \quad (31.1)$$

Since

$$E_k = I_p Z_k \quad (31.2)$$

the plate current is

$$I_p = \frac{\mu E_s}{r_p + Z_L + Z_k(1 + \mu)} \quad (31.3)$$

and the complex voltage amplification becomes

$$A_a = \frac{E_o}{E_s} = \frac{-\mu Z_L}{r_p + Z_L + Z_k(1 + \mu)} \quad (31.4)$$

Thus, the actual voltage amplification is decreased from

$$A = \frac{-\mu Z_L}{r_p + Z_L}$$

without cathode degeneration to the amount indicated by (31.4) with cathode degeneration.

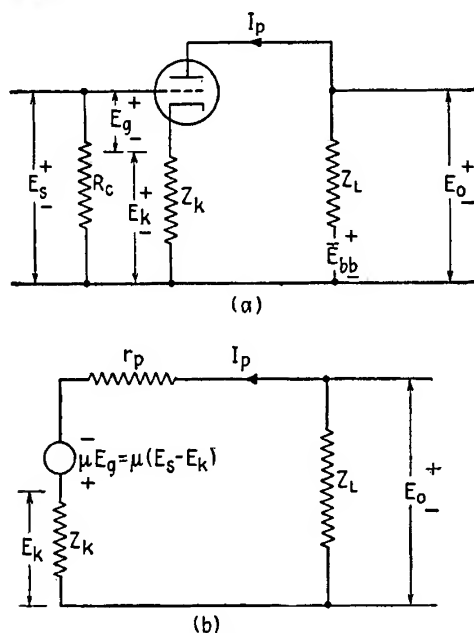


FIG. 31.1.—Cathode degeneration.

Inspection of (31.4) shows that, for no cathode degeneration to be present,

$$Z_k(1 + \mu) \ll r_p + Z_L$$

32. The Cathode Follower.—Figure 32.1a shows a simple “cathode-follower” circuit. There is no plate-load impedance, so that the potential of the plate never changes. The output voltage appearing across a cathode impedance Z_k “follows” closely the changes in potential of the grid, hence the name “cathode follower.” Since the voltage fed back is the voltage across the load,

the device is essentially a single-stage amplifier having voltage-controlled feedback with the feedback coefficient β equal to -1 . The cathode follower is characterized by low output impedance and high input impedance. Its voltage amplification is less than unity. Therefore the cathode follower is used, not for voltage amplification, but as a form of power amplifier with a low power

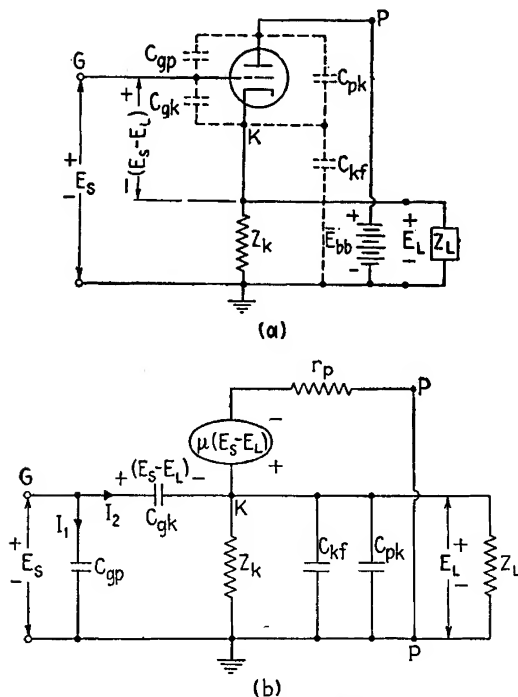


FIG. 32.1.—(a) Actual and (b) equivalent circuit diagram for a cathode-follower stage.

output. It is useful as a coupling device between a high-impedance source and a low-impedance load. An expression for the voltage amplification of a cathode follower may be obtained by use of the equivalent circuit of Fig. 32.1b. Neglecting tube electrode capacitances and introducing Z_L' as the parallel impedance of Z_k and Z_L ,

$$Z_L' = \frac{Z_k Z_L}{Z_k + Z_L} \quad (32.1)$$

the voltage relationship in the loop containing r_p and Z_L' yields

$$E_L = \mu(E_s - E_L) \frac{Z_L'}{r_p + Z_L'}$$

from which

$$A_a = \frac{E_L}{E_s} = \frac{\frac{\mu}{1+\mu} Z_L'}{\frac{r_p}{1+\mu} + Z_L'} \quad (32.2)$$

The same equation may be obtained from the feedback formula (29.3), if A is taken as the voltage amplification with the grid return lead, Fig. 32.1, moved from ground to cathode and β is taken as -1 . Thus

$$A_a = \frac{\frac{\mu Z_L'}{r_p + Z_L'}}{1 - (-1) \frac{\mu Z_L'}{r_p + Z_L'}}$$

which simplifies to (32.2). Equation (32.2) leads to the equivalent series circuit shown in Fig. 32.2 (the tube electrode capacitances

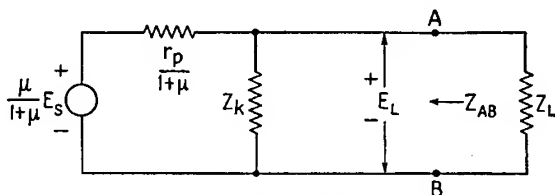


FIG. 32.2.—Simplified equivalent circuit diagram for a cathode-follower stage.

being neglected). From this circuit, the output impedance of the cathode follower is

$$Z_{AB} = \frac{\frac{r_p}{1+\mu} Z_k}{\frac{r_p}{1+\mu} + Z_k} = \frac{r_p Z_k}{r_p + (1+\mu) Z_k} \quad (32.3)$$

When a high- μ triode or pentode is employed, $\mu \gg 1$, (32.2) and (32.3) may be simplified to the forms

$$A_a \doteq \frac{Z_L'}{\frac{1}{g_m} + Z_L'} \quad (32.4)$$

$$Z_{AB} \doteq \frac{Z_k}{1 + g_m Z_k} \quad (32.5)$$

Sometimes a voltage-amplifier type of pentode having a large g_m is used with a cathode resistor of several thousand ohms and a

load having a high impedance, so that $g_m Z_k \gg 1$. Then, (32.4) and (32.5) simplify to

$$A_a \doteq 1 \quad (32.6)$$

$$Z_{AB} \doteq \frac{1}{g_m} \quad (32.7)$$

The effect of interelectrode capacitances may be accounted for as follows: Since the plate is at ground potential, C_{pk} is in parallel with Z_L together with C_{kf} , Fig. 32.1b. This will affect the value of Z_L' and A_a of (32.2). The input impedance, C_{gp} and C_{gk} being taken

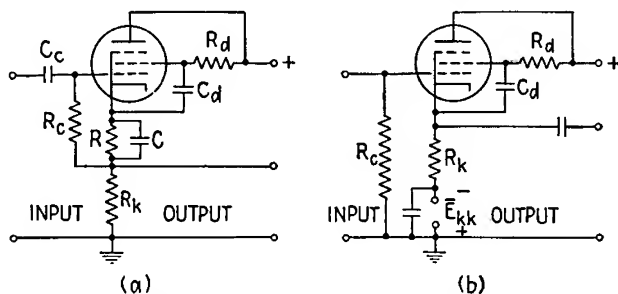


FIG. 32.3.—Practical cathode-follower circuits.

into account, may be determined readily by first considering the input admittance. The input admittance is determined by the currents I_1 and I_2 , Fig. 32.1b. These are given by

$$I_1 = j\omega C_{gp} E_s \quad (32.8)$$

$$I_2 = j\omega C_{gk} (E_s - E_L) \quad (32.9)$$

$$\text{Then } Y_{in} = \frac{I_1 + I_2}{E_s} = j\omega C_{gp} + j\omega C_{gk} (1 - A_a).$$

If A_a is real, Y_{in} is purely capacitive, and C_{in} is given by

$$C_{in} = C_{gp} + C_{gk} (1 - A_a) \quad (32.10)$$

If A_a is complex, Y_{in} has a real part, which may be either a positive or a negative conductance. Since A_a is often very nearly equal to unity, the input capacitance is often very small, only a few micromicrofarads.

The input capacitance of a cathode follower employing a screen-grid tube can be derived in a similar way, the exact form of the equation depending upon the particular way in which the screen grid is connected.

Since the d-c voltage across the cathode resistor is often of the order of several hundred volts, some arrangement for obtaining

proper grid-bias voltage is necessary. Two practical circuits are shown in Fig. 32.3. In Fig. 32.3a, the voltage drop across R alone is applied to the grid by means of the grid resistor R_c . Frequently the a-c voltage developed across R is negligible, and the by-pass capacitor C is omitted. In Fig. 32.3b, the proper bias voltage is obtained by means of a compensating voltage \bar{E}_{kk} introduced into the cathode lead from an external source.

33. Motorboating.—In the circuit of Fig. 33.1 the impedance Z_{bb} consists of the internal impedance of the plate power supply

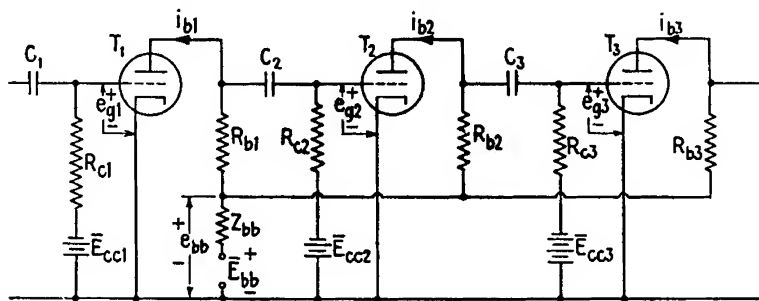


FIG. 33.1.—Feedback through the impedance of a common plate power supply.

and any added impedance that is in series with the plate power supply. Since Z_{bb} is common to the plate circuits of both tubes, it provides a feedback path.

Assume that a positive increment of potential is applied to the grid of T_1 . The total instantaneous value of plate current i_{b1} increases, thus increasing the potential developed across Z_{bb} and reducing the total instantaneous value of the effective plate power-supply potential e_{bb} . The resulting increase (in addition to the normal action of R_{b1}) in the negative value of e_{g2} results in a decrease in i_{b2} . This, in turn, tends to decrease the potential developed across Z_{bb} and to increase e_{bb} . Since owing to amplification the change in i_{b2} is usually greater than that in i_{b1} , the effective value of e_{bb} is increased, resulting in a decrease in the change of e_{g2} . This results in negative feedback. A similar effect occurs for any even number of stages in cascade.

When there are three stages, however, the increase in i_{b3} results in a decrease in e_{bb} since $|\Delta i_{b3}| > |\Delta i_{b2} - \Delta i_{b1}|$. The change in e_{bb} increases the change in e_{g2} , resulting in positive feedback, which may cause oscillation (cf. Chap. XV). A similar effect occurs for any odd number of stages.

If oscillation occurs, the grid potential of one of the tubes

usually goes positive, causing grid current. The resulting "grid-leak" biasing may block the amplifier and stop the oscillation. Since the time constant of the combination of the coupling capacitor and grid resistor is large, this results in a low-frequency oscillation that sounds very much like an outboard motor. Thus, this phenomenon is known as *motorboating*.

Motorboating may be overcome by (1) decreasing the low-frequency voltage amplification of the amplifier, (2) by using a well-regulated power supply, (3) by using decoupling networks as illustrated in Fig. 33.2 by R_{b1}' , C_{b1} ; R_{b2}' , C_{b2} ; and R_{b3}' , C_{b3} .

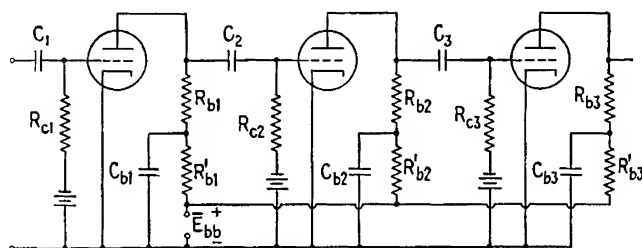


FIG. 33.2.—Decoupling networks to prevent motorboating.

Feedback due to the common plate impedance Z_{bb} is usually undesirable but may be employed to modify the frequency response of an amplifier by proper choice of Z_{bb} and the number of stages affected.

Supplementary Reading

- E. C. JORDAN, *et al.*, ed. by W. L. Everitt: "Fundamentals of Radio," pp. 186–189, Prentice-Hall, Inc., New York, 1942.
- J. B. HOAG: "Basic Radio," pp. 198–204, D. Van Nostrand Company, Inc., New York, 1943.
- R. S. GLASGOW: "Principles of Radio Engineering," pp. 212–220, McGraw-Hill Book Company, Inc., New York, 1936.
- W. L. EVERITT: "Communication Engineering," 2d ed., pp. 463–473, McGraw-Hill Book Company, Inc., New York, 1937.
- F. E. TERMAN: "Radio Engineering," pp. 248–256, McGraw-Hill Book Company, Inc., New York, 1937.
- MIT STAFF: "Applied Electronics," pp. 525–539, John Wiley & Sons, Inc., New York, 1943.
- K. HENNEY: "Radio Engineering Handbook," 3d ed., pp. 386–388, McGraw-Hill Book Company, Inc., New York, 1941.
- J. G. BRAINERD, *et al.*, eds.: "Ultra-high-frequency Techniques," pp. 104–119, D. Van Nostrand Company, Inc., New York, 1942.
- RCA: Receiving Tube Manual RC-14, pp. 20–22.
- H. J. REICH: "Theory and Application of Electron Tubes," 1st ed., pp. 220–223, 2d ed., pp. 197–211, McGraw-Hill Book Company, Inc., New York, 1944.

CHAPTER XIV

POWER TUBES

1. Definition of Power Tubes.—Although there is no sharp line of distinction between power tubes and other types, power tubes may be considered to be those tubes whose function it is to deliver a-c power to a load circuit. Power tubes described in this chapter are multielement tubes of the negatively biased control-grid type that are operated under such conditions as to make plate-circuit efficiency and dissipation at the electrodes important factors in design and operation. Though many power tubes are physically of large size, small tubes may be operated as power tubes. The power input to the plate circuit ranges from a few watts to several hundred kilowatts. The d-c plate voltage ranges from 100 volts or even less to 20 or 25 kilovolts. Several types of power tubes are shown in Figs. 1.1 to 1.3.

Large power tubes are commonly operated in the Class *B* or Class *C* modes to secure high output and high plate-circuit efficiency.

2. Structural Features of Power Tubes.—Before considering the operation of power tubes some general practical points concerning the structure of the tubes will be described. Many power tubes are triodes, though tetrodes and pentodes are often made in large sizes and used as power tubes.

Cathodes.—Although the coated cathode has the highest emission efficiency and is used almost universally in small tubes for receivers, it is generally impracticable for power tubes operated with high d-c plate voltages. The emission of coated cathodes may fail permanently or temporarily, and parts of the cathode coating may flake off, when high current is drawn for a considerable length of time. Coated cathodes are sometimes used when the plate voltage is 1,000 volts or less or in some special-purpose tubes when the plate current flows for very short intervals of time.

Thoriated-tungsten filaments, possessing higher emission efficiency than pure metal filaments, are generally used in power tubes when the d-c plate voltage is not in excess of 5,000 volts. The

delicate monatomic surface layer of thorium deteriorates under bombardment by positive ions from the residual gas.

Pure tungsten filamentary cathodes are used in large power tubes, especially those operated at very high plate voltages.

Although drawing any amount of emission current from tungsten filaments does not affect the emission efficiency or life of the filament, it is generally conceded that the activity of coated and thoriated cathodes gradually decreases if the average plate current exceeds some definite value determined by experiment. The decrease in life is caused largely by the presence of residual gas and the positive-ion bombardment resulting from it. Hence the loss in life and the necessity of limiting the average plate current depend upon the degree of vacuum that is maintained in the tube.

The life of a power tube is generally determined by the life of the cathode. Small power tubes have a life expectancy of about 2,000 hr, while large power tubes are operated at such temperatures as to give a much longer life, ranging up to 10,000 hr or more. Very large tubes are sometimes operated with the vacuum maintained by pumps so that burned-out filaments can be replaced.

Plates.—Most of the power loss within a tube is in heating the plate. The power rating of a tube is determined, therefore, by the amount of power the plate can get rid of, just as the power rating of a transformer or motor is determined by its temperature rise. Not only must the plate get rid of the heat developed directly by electron bombardment, but a fraction of the cathode-heating power that is radiated from the cathode and intercepted by the surrounding plate structure must be dissipated by the plate. Furthermore, a part of the power radiated from the grid is intercepted by the plate and hence contributes in small measure to the total power that the plate must dissipate.

The heat at the plate is dissipated most commonly either by radiation or by forced circulation of a cooling agent.

Since radiation depends upon the fourth power of the absolute temperature, the plate of a radiation-cooled tube is designed to operate at a high temperature and is mounted in a glass envelope through which the radiation must pass. If the plate is of tungsten or tantalum, it operates at a bright cherry red and may radiate about 8 watts/cm² of outside exposed surface. A molybdenum plate operates at a bright red and can dissipate about 5 watts/cm², while a nickel plate operates at a dull red and dissipates about 3 watts/cm² of surface. Often carbon plates are used because their

dull black surface makes them good radiators of heat. Their operating temperature is so low that there is no visible radiation. A radiation-cooled triode is shown in Fig. 1.1.

The plate of the large tubes is generally made of copper in tubular shape with one end closed and the other end sealed to glass, thus comprising a portion of the vacuum envelope. A water jacket around the plate permits cooling water to be forced rapidly over the outer surface of the plate. Large amounts of heat thus can be taken away from the plate. A typical water-cooled tube is shown in Fig.

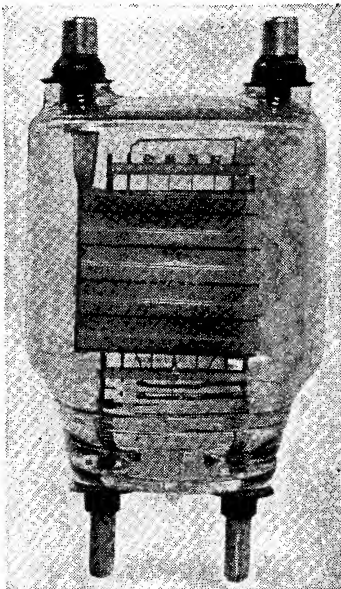


FIG. 1.1.—Radiation-cooled tube.

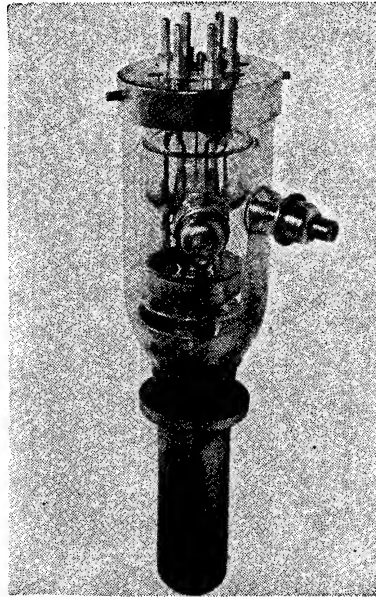


FIG. 1.2.—Water-cooled tube.

1.2. The tubular copper plate often is embedded in a block of copper provided with fins. Air forced between the fins cools the plate. These air-cooled tubes avoid the complicated water-cooling systems. An air-cooled tube is shown in Fig. 1.3.

Grids.—The grid of a power tube, located between plate and cathode which operate at high temperatures, receives considerable radiated heat from these electrodes. The added heat resulting from electron bombardment of the grid often elevates the temperature of the grid to such an extent that primary emission takes place from the grid wires. Such primary emission always causes a decrease in efficiency of the tube and may cause such excessive heating of the

plate as to destroy the tube. To withstand high temperatures the grid wires are made of tungsten, tantalum, or molybdenum. The grid wires get rid of heat generally by radiation assisted by conduction along the grid supports.

3. Heating of Filament.—Most power tubes have a filamentary cathode. The space current from the plate to the cathode (speaking in terms of the conventional direction of current flow) adds to or subtracts from the heating currents at the ends of the filament as indicated in Fig. 3.1. Consider a small part of the plate current,

Δi_b , which divides at the filament into two components $\Delta i_b'$ and $\Delta i_b''$. The sum of all the components $\Delta i_b'$ of all the current elements Δi_b to all parts of the filament flows against the heating current \bar{I}_f at the positive end of the filament, whereas the sum of all the $\Delta i_b''$ components adds to

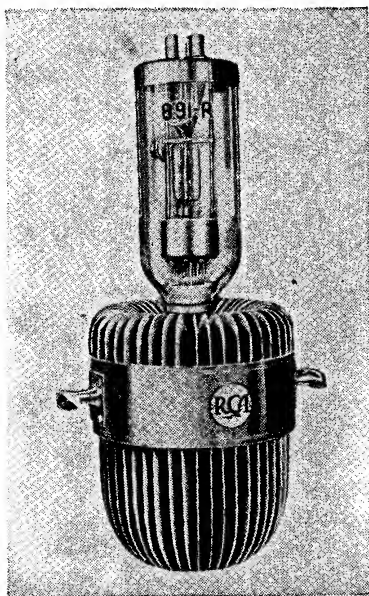


FIG. 1.3.—Air-cooled tube. (RCA.)

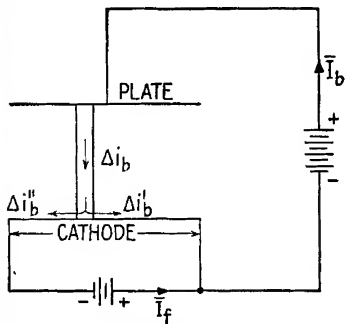


FIG. 3.1.—Division of space current in the filament.

the heating current at the negative end of the filament. Hence the negative end is overheated, while the positive end is underheated. In order to reduce this unequal heating of the filament, the heating current is generally from 10 to 30 times the average plate current.

Filaments of power tubes generally are heated by alternating current. In this case the heating of the filament, even as affected by the space current, is quite uniform. The connection of the plate circuit to the cathode circuit is generally made to the center point of the secondary winding of the heating transformers, Fig. 3.2a, or to the midtap of a shunting resistance, Fig. 3.2b. These

connections reduce the variations of actual plate-to-cathode potential caused by the voltage drop along the cathode. Capacitors, as shown in the figures, provide by-passes for the alternating components of the plate current.

The proper temperature of the filament is indicated preferably by a certain reading of the filament voltmeter rather than by a reading of an ammeter giving the filament current; for as the cross section of the filament decreases with age, maintaining a constant current would cause the temperature of the filament to increase.

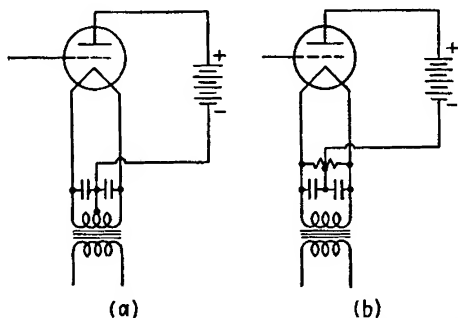


FIG. 3.2.—Plate-circuit connections to the cathode heating circuit.

4. High-frequency Tubes.—When tubes are operated at very high frequencies, the charging current into the interelectrode capacitances becomes very large, necessitating large lead-in wires. The interelectrode capacitances, especially that between grid and plate, are reduced to a minimum by bringing the leads out through the glass by the shortest path instead of running them all through a common base. The inductance of the leads also is reduced by making all current-carrying connections of large cross section within the tube.

Since losses in insulating material exposed to high-frequency fields increase with frequency, insulating supports for the grids and plate are largely eliminated, the electrodes often being supported by the seal-in leads.

The principal reason for the decrease in efficiency of a triode or tetrode as the frequency is increased is the appreciable time, compared with a period of oscillation, required for an electron to pass from cathode to plate. This *transit time* introduces a time lag between the change of plate current with respect to change of plate voltage; and, as will be pointed out later, any such phase difference decreases the efficiency. Furthermore, some electrons

that are attracted by the positively charged grid or plate may be pushed back to the cathode when the grid or plate potential reverses before these electrons have time to reach the grid or plate. In such a case, work will have been done on these electrons, and this work will be expended on the cathode when they return to it, overheating the cathode and causing useless power loss. In order to minimize this effect, the distances between electrodes may be decreased and the electrode voltages increased. For each type of tube there is a more or less definite upper frequency limit above



FIG. 4.1.—Type 211 tube.

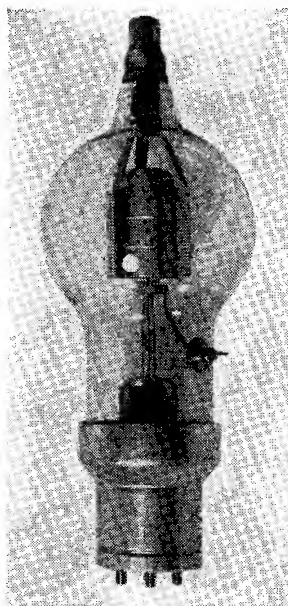


FIG. 4.2.—Radiation-cooled tube type 454.

which the efficiency of the tube falls off rapidly. For example, the type 211 tube, Fig. 4.1, has all lead-in wires pass through a common base. It is designed primarily for audio or broadcast frequencies and works at full output up to about 15 mcps. On the other hand, the type 454, Fig. 4.2, is designed for high frequencies and operates satisfactorily at frequencies as high as 150 mcps.

5. Classification.—Power tubes are classified conveniently in two groups according to whether they operate within a narrow band of frequencies or through a wide band of frequencies. This classification is given below with the various applications listed in each group.

I. Narrow-band operation:

1. Amplifiers, Classes *B* and *C*
 - a. Modulated
 - b. Unmodulated
2. Frequency multipliers, Class *C*
3. Oscillators, Class *C*
 - a. Modulated
 - b. Unmodulated

II. Wide-band operation:

1. Audio amplifier, Class *B*
2. Modulators, Classes *B* and *C*

The tubes in Group I operate usually into a tuned plate load and are excited at the grid by a pure or modulated sinusoidal voltage. These tubes may be called *sine-wave power converters*. The tubes of Group II operate into plate loads that offer an appreciable impedance at all frequencies over a wide range, and their grids are excited by voltage waves containing many frequency components. These tubes may be called *nonsinusoidal power converters*. The operation of the Class *B* audio amplifier is described in Chap. XIII. Modulator tubes include not only tubes used to speech-modulate an amplifier or oscillator but also various types of switching, pulsing, and keying modulators.

6. General Principle of Operation of a Power-converter Tube.—

The mode of operation of a power-converter tube is generally so different from that of a Class *A* amplifier that little is gained in trying to maintain an analogy. The Class *A* tube is usually a linear device, while the power-converter tube is entirely nonlinear. This nonlinear action inevitably results from applying large voltage variations on the electrodes of the tube in order to obtain reasonable powers at satisfactory efficiencies.

The object of the power converter is to draw d-c power from the plate-circuit source \bar{E}_{bb} , Fig. 6.1, and to convert that power into a-c power in the load. This is accomplished by causing properly timed current pulses to flow through the plate load much as a controlled switch might do, as indicated in Fig. 6.2. The switch would be closed periodically for short intervals of time. The power tube may be considered as a sort of switch controlled by the grid voltage.

When the switch in Fig. 6.2 is closed, the full voltage \bar{E}_{bb} acts across the load. If the resistance of the switch is zero and the

switch could be opened and closed with no sparking, the power conversion would be effected at 100 per cent efficiency (actually, sparking could not be avoided). The "vacuum-tube switch" offers some resistance, and therefore there is a voltage across the tube while the current flows. The *smaller* the voltage e_b across the tube while the current flows, the *lower* the *power loss* in the tube and hence the *higher* the *efficiency* of power conversion.

The grid-bias voltage \bar{E}_{cc} polarizes the grid negatively to cut off the plate current, *i.e.*, to hold the switch open, except when the grid input voltage e_g periodically overcomes the negative bias and causes current to flow.

The operation of the power converter can be described qualitatively by the steady-state voltage and current waveforms. Sup-

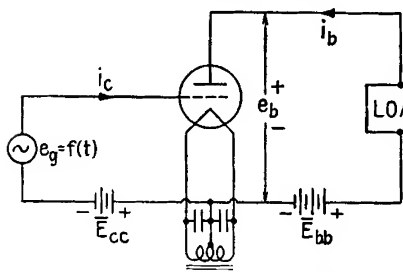


FIG. 6.1.—Power-converter tube.

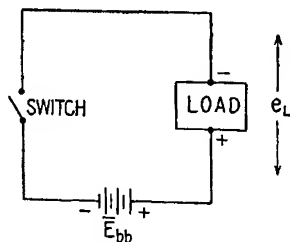


FIG. 6.2.—Equivalent circuit of a power-converter tube.

pose, for example, that the plate load is a resistor having a pure resistance of constant value. If the grid-excitation voltage is sinusoidal, Fig. 6.3a shows the waveforms of e_c , e_b , and i_b . Since i_b depends upon the equivalent voltage $e_c + e_b/\mu$, plate current flows only when this equivalent voltage is positive and hence only between the instants when $e_c = -e_b/\mu$. This value of e_c , which just makes i_b equal to zero, is called the cutoff grid voltage. The value of cutoff e_c can be found easily by plotting $-e_b/\mu$ as shown by the dashed line in Fig. 6.3a. The plate current starts when the grid voltage, increasing in the positive direction, reaches cutoff. It should be noted in Fig. 6.3a that the peak of the plate current i_b occurs when e_b is a minimum. This (as previously explained) is the condition for high efficiency and low plate loss. In Fig. 6.3b the waveforms are shown for a square-wave grid voltage. In this particular case the ratio of average power fed to the load to that lost in the tube is $[\bar{E}_{bb} - (e_b)_{\min}]/(e_b)_{\min}$, and the efficiency of plate-circuit conversion is $[\bar{E}_{bb} - (e_b)_{\min}]/\bar{E}_{bb}$. Hence the smaller

$(e_b)_{\min}$, the higher the efficiency. For Fig. 6.3a the calculation of powers involves integrations, but it is evident by analogy that, the lower $(e_b)_{\min}$, the higher the efficiency.

If the plate load is a tuned circuit (commonly called a "tank circuit"), Fig. 6.4, the system amplifies only within the narrow

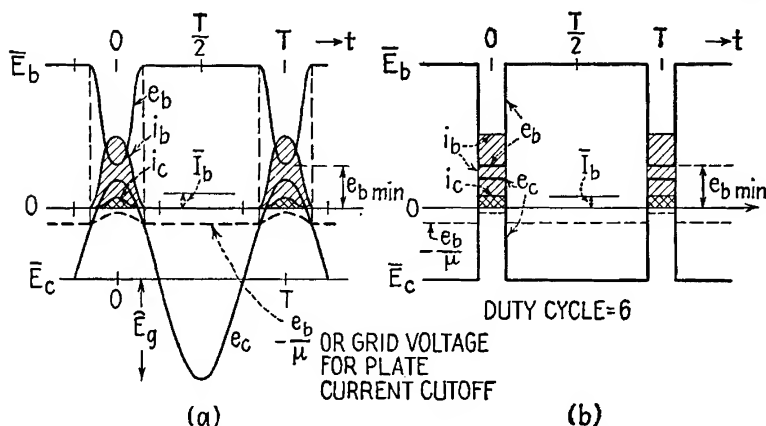


FIG. 6.3.—Current and voltage waveforms for a power-converter tube having a resistor as a plate load.

frequency band for which the plate load offers an appreciable impedance. If the plate load is tuned for parallel resonance at the frequency of e_g , the plate load offers a pure-resistance load of value $R_L = L_b/C_b R_b$. The waveforms are shown in Fig. 6.5a.

The duration of the plate-current pulse can be found as for Fig. 6.3, by plotting $-e_b/\mu$. So long as e_c is negative and in magnitude greater than e_b/μ , the plate current is zero. This is true from the beginning of the half wave at $-T/4$ to $-t_b$, Fig. 6.5a. The plate current then rises to a maximum, which occurs when the plate voltage e_b is a minimum, and then falls to zero at t_b before the end of the half cycle. The grid current flows only while the grid voltage is positive and hence for a time less than $2t_b$, Fig. 6.5a.

If the grid-bias voltage is greater, the plate current flows in shorter pulses, Fig. 6.5b. A larger fraction of the plate current

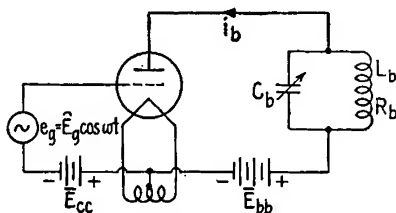


FIG. 6.4.—Power-converter tube with a tuned plate load, for narrow-band operation.

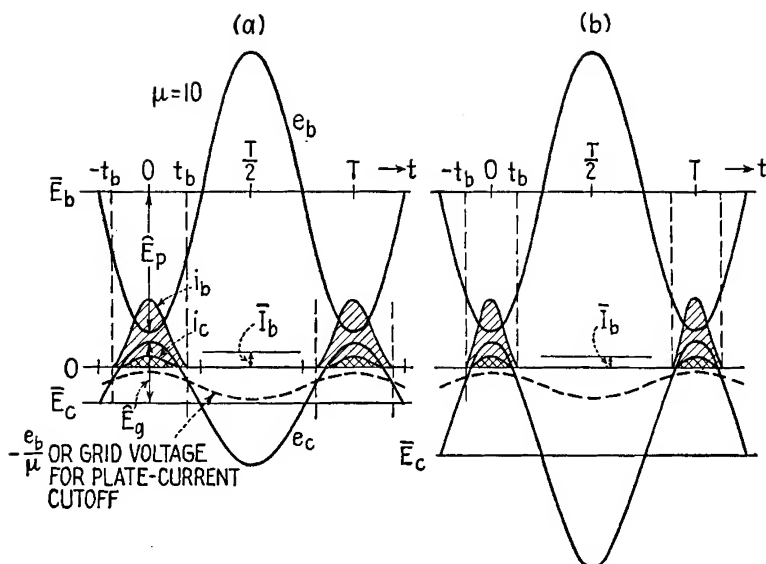


FIG. 6.5.—Current and voltage waveforms for a power-converter tube having a tuned plate load; (a) angle of plate-current flow approximately 130° ; (b) a smaller angle of plate-current flow.

flows while the plate voltage is low, giving a higher efficiency of power conversion.

Suppose now that the plate circuit is detuned so that the plate-

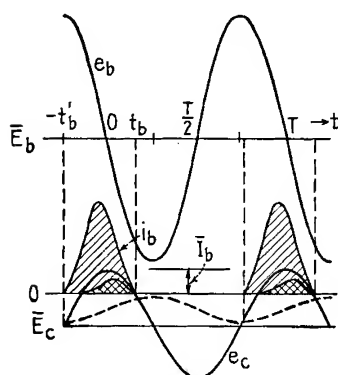


FIG. 6.6.—Same as Fig. 6.5 but with a reactive load.

voltage minimum does not occur at the same instant that e_c is a maximum. This condition is represented in Fig. 6.6. The plate-current waveform is largely determined by the grid-voltage waveform but the plate-current pulse is shifted slightly to the left, Fig. 6.6, because the plate voltage is larger to the left of the peak of the grid-voltage wave than to the right. Hence the plate current flows when the plate voltage is higher than $(e_b)_{\min}$, and the plate loss consequently is much increased, resulting in a decrease in efficiency.

Also, each pulse of plate current is larger owing to the larger values of e_b during the pulse. The average plate current and the d-c power input increase, causing a decrease in efficiency. Hence for maximum

efficiency it is essential that the plate and grid voltages be *opposite in phase*.

The shift of the alternating component of the plate voltage to the right, Fig. 6.6, causes the grid-current pulse to shift to the right with relation to the grid-voltage wave, because, the lower the plate voltage, the higher the grid current.

The conclusions drawn from the preceding discussion may be summarized as follows: *The plate-circuit efficiency of a power tube in narrow-band operation is (1) greatest when the alternating plate and grid voltages are opposite in phase (2) is increased by decreasing the minimum plate voltage, and (3) is increased by decreasing the duration of the plate-current pulse.*

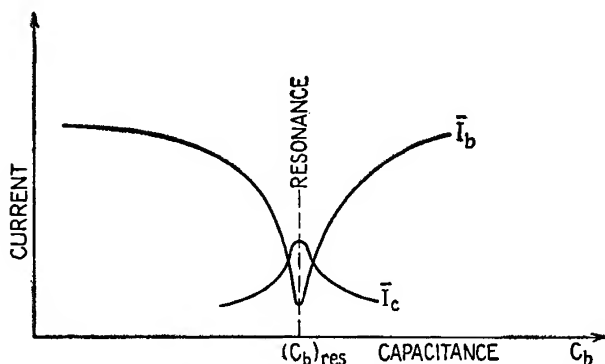


FIG. 6.7.—Effect of detuning the plate tank circuit upon the average plate and grid currents.

The condition of proper phase is obtained by tuning the plate-circuit load to unity power factor as indicated by a *minimum* value of average plate current \bar{I}_b . While the average plate current is a minimum at tank-circuit resonance, the average grid current is a maximum, Fig. 6.7.

Lower minimum plate voltages are secured by increasing R_L . The plate-current pulses are shortened by increasing the negative grid bias.

There are other factors, to be explained later, that may be of more importance than plate-circuit efficiency and that must be considered in choosing the conditions of operation of the system.

7. Quantitative Relationships in Narrow-band Operation.—In order to understand fully the power-converter tube and to be able to design and to adjust a system it is necessary to study the converter from a quantitative point of view. The purpose in this section is to

list and define the various quantities required for the subsequent analysis.

The circuit diagram of a narrow-band power amplifier is shown with series feed of the d-c power in Fig. 7.1. This connection has the practical disadvantage that both plate and grid tank circuits

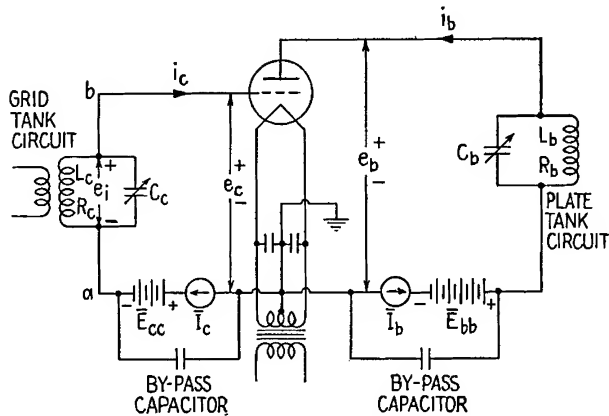


FIG. 7.1.—Series-feed connection of a power amplifier.

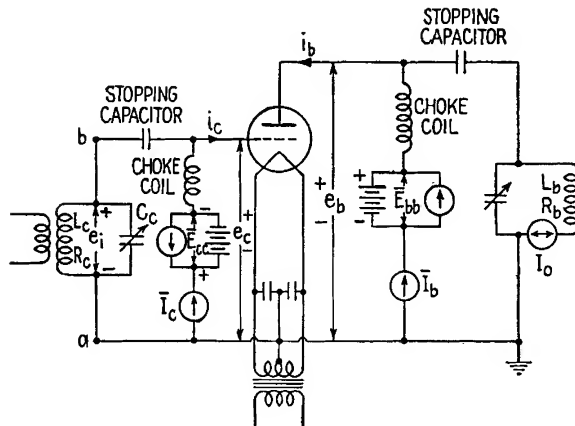


FIG. 7.2.—Parallel-feed connections of a power amplifier.

are maintained at steady voltages above and below ground, of amounts \bar{E}_{bb} and \bar{E}_{cc} respectively. The alternative parallel-feed connections of Fig. 7.2 eliminate this disadvantage. The two circuits are electrically equivalent provided that each of the choke coils has sufficient impedance to eliminate most of the alternating currents from the d-c paths. In practice, it is advisable to connect

protecting by-pass capacitors across the power sources and the d-c ammeters in each circuit. It should be noted that, although tuned circuits are shown in the grid circuits of Figs. 7.1 and 7.2, these circuits are not considered as a part of the stage being studied. The sinusoidal voltage $e_i = \hat{E}_{ci} \cos \omega t$ is assumed to act between points a and b , and the circuits from which e_i is obtained are properly a part of the previous stage, or driving source.

The various voltages and currents are indicated on the diagrams. The alternating components of the grid and plate voltages are assumed sinusoidal in form and in phase opposition, in order to conform to the best operating conditions and to be given by¹

$$e_c = \bar{E}_c + \hat{E}_g \cos \omega t \quad (7.1)$$

$$e_b = \bar{E}_b - \hat{E}_p \cos \omega t \quad (7.2)$$

Since the circuits associated with the tube usually have little d-c resistance, the average grid- and plate-polarizing voltages \bar{E}_c and \bar{E}_b are practically equal in magnitude to the supply voltages, or

$$\bar{E}_c \doteq -\bar{E}_{cc} \quad (7.3)$$

$$\bar{E}_b \doteq \bar{E}_{bb} \quad (7.4)$$

Furthermore, the grid voltage \hat{E}_g is practically the same as the input or driving voltage \hat{E}_i , applied to the terminals ab , Figs. 7.1 and 7.2, and the plate voltage \hat{E}_p is practically equal in magnitude to the voltage across the load.

If the frequency is low enough so that the electron transit time from cathode to plate is inappreciable compared with the period of oscillation, i_c and i_b are determined solely by e_b and e_c . Since the tube is a nonlinear device, the grid- and plate-current variations are nonsinusoidal in response to the sinusoidal variations of grid and plate voltages, as shown in Figs. 6.5*a*, *b*, and can be expressed by the Fourier series

$$i_c = \bar{I}_c + (\hat{I}_g)_1 \cos \omega t + (\hat{I}_g)_2 \cos 2\omega t + \dots \quad (7.5)$$

$$i_b = \bar{I}_b + (\hat{I}_p)_1 \cos \omega t + (\hat{I}_p)_2 \cos 2\omega t + \dots \quad (7.6)$$

In a practical system \bar{E}_{cc} and \bar{E}_{bb} are indicated by d-c voltmeters. The average, or d-c, grid and plate currents are indicated by d-c ammeters connected as in Figs. 7.1 and 7.2.

The various power quantities that now will be defined are, for

¹ Where amplitude values are used in this chapter, magnitudes only are assumed, for example, \hat{E}_g and \hat{E}_p in (7.1) and (7.2).

any given tube, dependent in magnitude solely upon the value of the four voltages \bar{E}_c , \bar{E}_b , \hat{E}_p , \hat{E}_g .

In the plate circuit, the *power to the plate circuit* from the d-c power source is

$$P_{bb} = \bar{E}_{bb}\bar{I}_b \doteq \bar{E}_b\bar{I}_b \quad (7.7)$$

The *power delivered to the plate load* depends only upon the a-c component of (7.2) and the fundamental component of (7.6) and is

$$P_L = \frac{\hat{E}_p(\hat{I}_p)_1}{2} = \frac{\hat{E}_p^2}{2R_L} \quad (7.8)$$

The power input less the power delivered to the load is the *plate dissipation*, or

$$P_p = P_{bb} - P_L \quad (7.9)$$

The *plate-circuit efficiency* is the ratio of a-c power delivered to the load to the d-c power input to the plate circuit, or

$$\eta_p = \frac{P_L}{P_{bb}} \quad (7.10)$$

In the grid circuit, the generator or source from which the signal voltage is derived supplies an input or *driving power* equal to

$$P_d = \frac{\hat{E}_g(\hat{I}_g)_1}{2} \quad (7.11)$$

The d-c grid current flows *into* the positive terminal of the d-c grid-voltage supply; hence power is delivered to rather than taken from the grid-bias supply. If the grid bias is furnished by a battery, this battery is charged by the grid current of the tube. The *power supplied to the grid-bias source* is

$$P_{cc} = \bar{E}_{cc}\bar{I}_c \quad (7.12)$$

and must be obtained from the input or driving power P_d supplied by the grid driving source. The difference between the grid-circuit input power and the power supplied to the bias source is dissipated at the grid. The *grid dissipation* is

$$P_g = P_d - P_{cc} \quad (7.13)$$

The *power amplification* A_p is the ratio of the a-c power delivered to the plate load to the input or grid driving power, or

$$A_p = \frac{P_L}{P_d} \quad (7.14)$$

Equations (7.9) and (7.13) are true only when there is no power transfer in either direction between grid and plate circuits. Power interchange between these circuits can take place in two ways, (1) through any coupling between circuits such as that provided by capacitance between grid and plate electrodes and wiring and (2) by the passage of electrons directly from grid to plate, or vice versa. This second effect takes place when secondary emission occurs at one or the other electrode. The effects of secondary emission in power tubes will be discussed in Sec. 9. The first cause of power interchange can be reduced or eliminated by neutralization, Sec. 13.

8. Testing and Operation of Power Tubes.—The various components of power that are important in the operation of a power tube used as a power converter have been listed in Sec. 7. Under test or under operating conditions these powers may be measured in several ways.

The values of P_{bb} and P_{cc} always can be determined from the readings of d-c instruments. The power P_L is the total a-c power delivered to the complete circuit external to the tube. The useful fraction of P_L depends upon the efficiency of the plate-load circuits; but, in considering the efficiency of the tube itself, P_L is considered to be the total output power from the tube. This output power is given approximately by the expression $I_o^2 R_b$ where R_b is the total effective series resistance of the tank circuit and I_o is the alternating current in R_b , Fig. 7.2. Sometimes P_L is obtained by measuring the heat developed in a given time in the load resistance or by matching the light given out by incandescent lamps used as load resistance to the light in the same or similar lamps given by a measured d-c or a-c power of low frequency.

From measured values of P_{bb} and P_L , P_p can be determined from (7.9). In some cases, especially when the plate of a tube is water-cooled, P_p can be obtained by measuring the rate at which the heat is carried away by the cooling water. Then (7.9) can be used for determining P_L .

In the grid circuit, P_d is given accurately by

$$P_d = \frac{1}{T} \int_0^T e_i i_c dt \quad (8.1)$$

Since e_i is assumed as $\hat{E}_g \cos \omega t$, the varying component of (7.1),

$$P_d = \frac{\hat{E}_g}{T} \int_0^T i_c \cos \omega t dt \quad (8.2)$$

Upon substituting for i_c its value given by (7.5), the integration yields the result given in (7.11). Since $(\hat{I}_g)_1$ cannot be measured easily, the expression (8.2) for P_d is not easily evaluated from instrument readings. There is an approximate method¹ for calculating P_d from measurements of \hat{E}_g and \bar{I}_c . Assume that i_c flows principally when $\cos \omega t$ has the value unity, *i.e.*, when e_c has its peak value or when e_g has its maximum positive value. That this is not strictly true can be seen from Fig. 6.5, but the assumption gives results that are sufficiently close to the true results to be useful. If $\cos \omega t$ in (8.2) be replaced by unity,

$$P_d \doteq \hat{E}_g \frac{1}{T} \int_0^T i_c dt = \hat{E}_g \bar{I}_c \quad (8.3)$$

Since $\cos \omega t$ in (8.2) never can exceed unity and has values less than unity some of the time, (8.3) evidently gives too large a result although the error may not exceed a few per cent.

9. Static Characteristic Curves of Power Tubes.—Calculation from the static characteristic curves is a practical method of determining the capabilities of a tube and also is convenient in the design of the circuits. It is especially valuable for large tubes because it avoids the difficulty and expense of setting up the tube and testing it. This section will be devoted to an explanation of the static curves of a power tube. A later section will deal with the method of calculation of the operating characteristics based upon the static curves.

The static characteristic curves that are most adaptable to the determination of power-tube performance are the constant-plate current and constant-grid current curves plotted to plate and grid voltages, Fig. 9.1. This particular tube, type 211, has a maximum plate dissipation of 100 watts and an amplification factor of 12. The curves for constant plate current are nearly straight and parallel wherever e_b is greater than about $2e_c$. Below the radial line for e_b equal to $2e_c$ the plate-current lines bend to the right and become nearly parallel to the e_c axis when e_c is greater than about $3e_b$. These factors of 2 and 3 vary according to the value of μ for the tube and therefore are not to be considered as important or critical factors. The reason for the sharp bending of the plate-current lines between the two radial lines just mentioned is the "stealing" of current from the plate by the grid as the ratio e_c/e_b increases. The

¹ H. P. THOMAS, Determination of Grid Driving Power in Radio Frequency Power Amplifiers, *Proc. I.R.E.*, **21**, 1134, 1933.

negative of the slope of the plate-current curves is the value of μ at that point. The value of μ decreases to zero for large values of e_c/e_b .

The curves for constant grid current are shown by the dashed lines. They fan out in a regular manner except for a region that for this tube is above the radial line $e_b = 2e_c$. The irregularity here is caused by secondary electron emission from the grid to the plate. The electrons that strike the grid may release from the grid

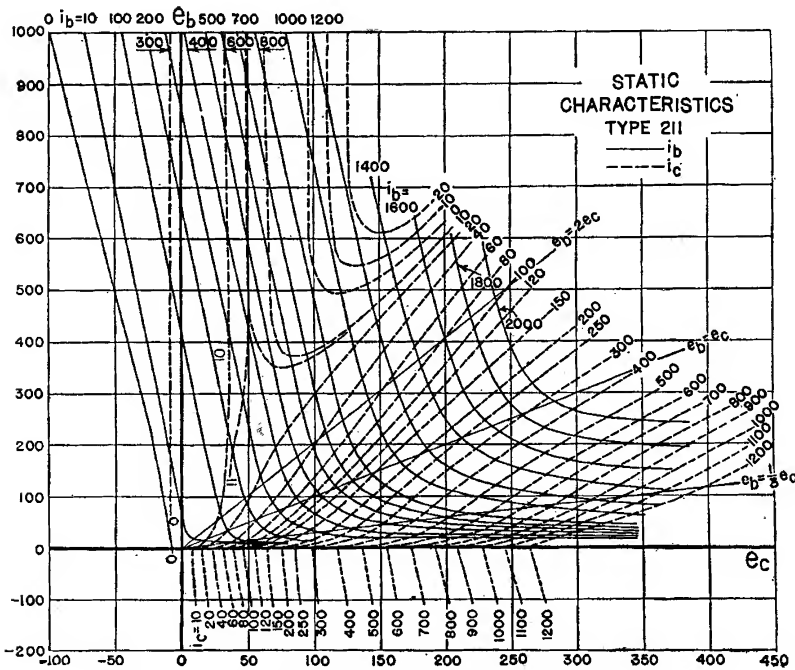


FIG. 9.1.—Static characteristic curves of the type 211 power tube.

a number of secondary electrons dependent upon the character of the surface of the grid. If the grid is "contaminated" by a layer of active material (such as thorium or barium) that has evaporated from the cathode, one bombarding electron may release an average of more than one secondary electron. The secondary electrons are attracted to the plate if e_b is greater than e_c , resulting in a decrease in grid current and an increase in plate current. If the number of secondary electrons released exceeds the number of primary electrons that strike the grid, the grid current is reversed. For example, Fig. 9.1, i_c is -20 ma when e_b and e_c are 610 and 150 volts,

respectively. If the grid is uncontaminated, the grid current generally is affected little by secondary emission. The grid-current lines then are spaced regularly for large values of e_b/e_c , as they are also for small values of the ratio. The effect of negative grid currents upon the operation of the tube will be discussed later.

The static characteristic curves often are provided by the manufacturer of the tube. They cannot be determined by observing the reading of d-c instruments as is the usual procedure for small receiving tubes because of the excessive heating of the electrodes. This is apparent by referring to Fig. 9.1 and calculating P_p for the point $e_b = 400$ volts and $e_c = 200$ volts. The plate current is 1.6 amperes giving a value of P_p of 640 watts. This is 6.4 times the maximum allowable value of P_p . The value of P_g at this point is $0.095 \cdot 200 = 19$ watts. This value of P_g is not greatly in excess of the maximum permissible value; but if e_b is 100 volts and e_c is 300 volts, P_g is $0.840 \cdot 300$, or 252 watts, a power that would melt the grid.

There are various special methods of obtaining the static curves in which the potentials are applied for short intervals of time while the measurements are made, thus reducing the average power losses at the electrodes.¹ Another method of obtaining the static curves is to extrapolate curves obtained for low values of voltages that can be determined by the simple d-c method.²

10. Path of Operation.—The path of operation of a Class A amplifier is drawn most conveniently on the i_b - e_b chart of the characteristic curves. This chart, however, is not convenient for power-tube operation because on the i_b - e_b chart the path of operation for a power tube is curved and difficult to determine. However, on the e_b - e_c chart, Fig. 9.1, the path for a power tube has a very simple form and is derived as follows.

The Q point is located first, as in Fig. 10.1, where the curves of Fig. 9.1 are partly reproduced. The value of \bar{E}_b , which is generally the same as \bar{E}_{bb} , is located on the e_b axis. The value of \bar{E}_c , which is generally equal to $-\bar{E}_{cc}$, then is located on the e_c axis. These two coordinates determine the position of the Q point, Fig. 10.1.

¹ E. L. CHAFFEE, Power Tube Characteristics, *Electronics*, **11**, 34, June, 1938; H. N. KOZANOWSKI and I. E. MOUROMTSEFF, Vacuum Tube Characteristics in the Positive Grid Region by an Oscillographic Method, *Proc. I.R.E.*, **21**, 1082, 1933.

² E. L. CHAFFEE, Characteristic Curves of a Triode, *Proc. I.R.E.*, **30**, 383, 1942.

The waveforms of e_c and e_b are given by (7.1) and (7.2) and are plotted in Fig. 10.1. At $t = 0$ the position of the point of operation is obtained by measuring downward from Q the voltage \hat{E}_p and to the right of Q the voltage \hat{E}_g . This locates the end point p of the path of operation, corresponding to the point of maximum instantaneous grid voltage. The instantaneous plate voltage at this time is a minimum and is equal to $\bar{E}_b - \hat{E}_p$, and the instantaneous grid voltage is a maximum and is equal to $\bar{E}_c + \hat{E}_g$. Other

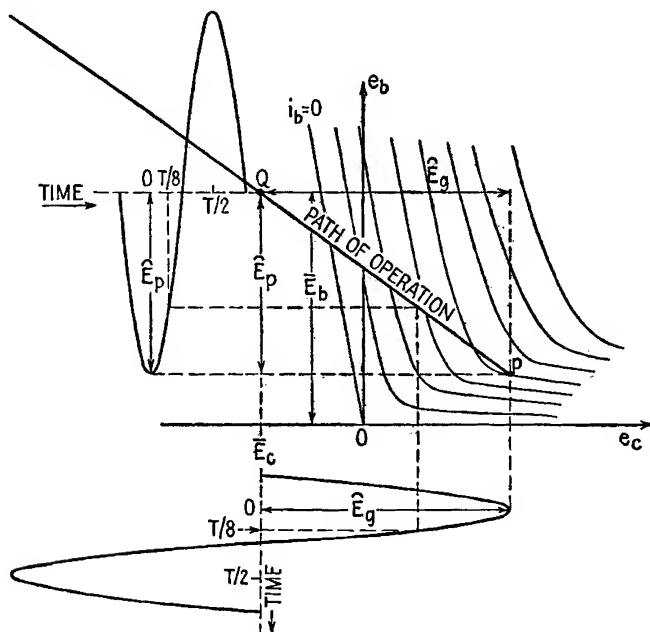


FIG. 10.1.—Path of operation when the alternating voltages on plate and grid are sinusoidal and in phase opposition.

positions of the operating point, as for example when $t = T/8$, lie on a *straight line* drawn through Q and p , Fig. 10.1.

The straight-line path is obtained only when the alternating plate and grid voltage are *sinusoidal* and in *phase opposition*. Since, as was explained in Sec. 6, the minimum plate loss results when these voltages are opposite in phase, the straight-line path is considered to be the normal type of path. If the phase of the two voltages departs from the 180-degree relation, the path opens into an ellipse, Fig. 10.2a. Harmonic voltages in either the grid or the plate voltage cause distortion of the path. For example, a

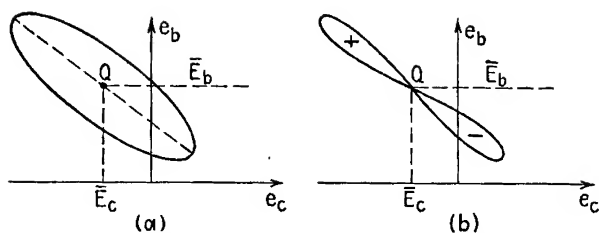


FIG. 10.2.—(a) Path of operation when E_p and E_g are not opposite in phase; (b) path of operation when E_p and E_g are opposite in phase, but e_p contains a second-harmonic component.

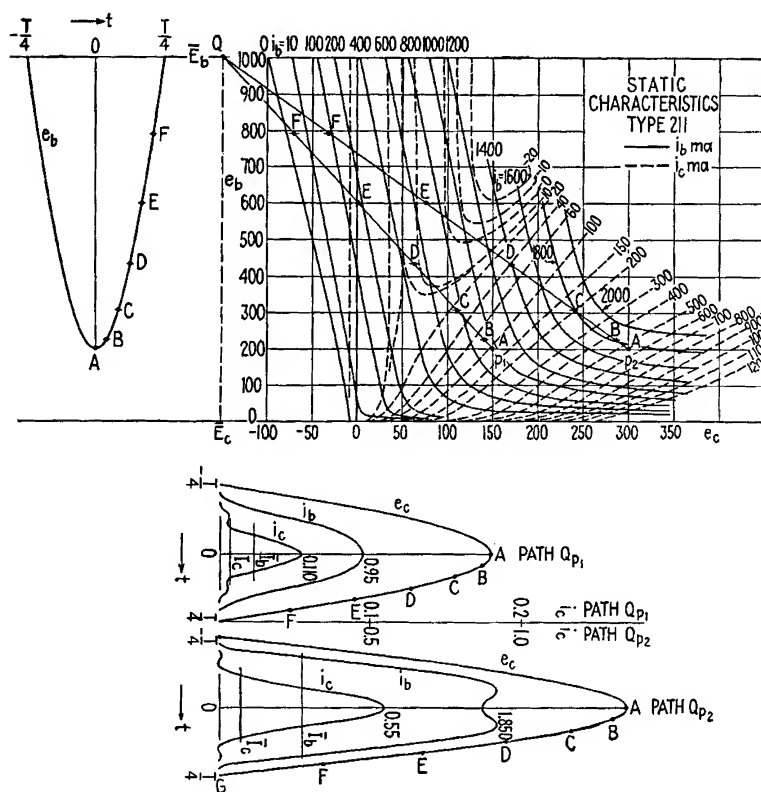


FIG. 10.3.—Current and voltage waveforms for two paths of operation.

second harmonic in the plate voltage causes the path to have the form of a figure 8, Fig. 10.2b. Although this is an open figure, the fundamental components are opposite in phase provided that the sum of the areas included within the path is zero, adjacent areas being given opposite signs. High-frequency components

of voltage may be present when the tube is overdriven or the circuits are improperly proportioned. Such high-frequency components cause other irregularities in the path.

Having drawn a path of operation as determined by the four given voltages \bar{E}_b , \bar{E}_c , \bar{E}_p , \bar{E}_g , the waveforms of the plate and grid currents can be determined easily. For example, two paths Qp_1 , Qp_2 , are drawn on the static characteristic curves of a particular tube, Fig. 10.3. The values of the four voltages for the paths are

Path, Fig. 10.3	\bar{E}_b	\bar{E}_c	\bar{E}_p	\bar{E}_g
Qp_1	1,000	-150	800	300
Qp_2	1,000	-150	800	450

The waveforms of the plate and grid currents, as derived from these paths, are drawn in the figure on the time axis for the grid-voltage waveform. The grid- and plate-voltage curves are shown for only the half cycle in which grid and plate currents flow. Note that the scale for i_b is the same for both paths but that the scale for i_c is different for the two paths and different from the scale for i_b .

11. Calculation of the Operation Condition.—Having obtained the waveforms of the plate and grid currents it is possible by Fourier analysis to find the average values \bar{I}_b and \bar{I}_c of the currents and the amplitudes of the fundamental components \hat{I}_{p1} and \hat{I}_{g1} . With these values and the known voltages \bar{E}_b , \bar{E}_p , \bar{E}_c , \bar{E}_g , given in the previous section for the paths of operation shown in Fig. 10.3, all the quantities given in (7.7) through (7.14) can be calculated.

A special harmonic analysis¹ developed for use in power-tube calculations enables the harmonic components of the plate and grid currents to be calculated directly from the path of operation. First, distances from Q along Qp are laid off, which are the following fractions of the length Qp 1, 0.966, 0.866, 0.707, 0.500, and 0.259. These points are denoted by the letters A through F as shown in Fig. 10.3 for the two paths Qp_1 and Qp_2 . The values of the instantaneous current at these points are read from the characteristic curves and are denoted by i_A , i_B , . . . , i_F . Then the average and fundamental components of the current are given by the formulas

¹ Chap. IX, Secs. 16, 17.

$$\bar{I} = \frac{1}{12} \left[\frac{i_A}{2} + i_B + i_C + i_D + i_E + i_F \right] \quad (11.1)$$

$$(\hat{I})_1 = \frac{1}{12} [i_A + 1.93i_B + 1.73i_C + 1.41i_D + i_E + 0.52i_F] \quad (11.2)$$

A simpler analysis, making use of only the A , C , and E points, gives the average and fundamental components by the formulas

$$\bar{I} = \frac{1}{6} \left[\frac{i_A}{2} + i_C + i_E \right] \quad (11.3)$$

$$(\hat{I})_1 = \frac{1}{6} [i_A + 1.73i_C + i_E] \quad (11.4)$$

Generally, these latter formulas are sufficiently accurate for power-tube calculations.

The average and fundamental components of both the plate and grid currents are calculated below for path Qp_1 , the more accurate formulas given in (11.1) and (11.2) being used.

$$\begin{aligned} \bar{I}_b &= \frac{1}{12} \left[\frac{0.960}{2} + 0.920 + 0.740 + 0.500 + 0.180 + 0.003 \right] \\ &= 0.236 \text{ ampere} \end{aligned}$$

$$\begin{aligned} (\hat{I}_p)_1 &= \frac{1}{12} [0.960 + 1.93 \cdot 0.920 + 1.73 \cdot 0.740 \\ &\quad + 1.41 \cdot 0.500 + 0.180 + 0.52 \cdot 0.003] = 0.402 \text{ ampere} \end{aligned}$$

$$\begin{aligned} \bar{I}_c &= \frac{1}{12} \left[\frac{0.110}{2} + 0.074 + 0.025 + 0.0103 + 0.001 + 0 \right] \\ &= 0.014 \text{ ampere} \end{aligned}$$

$$\begin{aligned} (\hat{I}_g)_1 &= \frac{1}{12} [0.110 + 1.93 \cdot 0.074 + 1.73 \cdot 0.025 \\ &\quad + 1.41 \cdot 0.0103 + 0.001 + 0] = 0.026 \text{ ampere} \end{aligned}$$

The same components as calculated by the simpler formulas (11.3) and (11.4) are

$$\bar{I}_b = \frac{1}{6} \left[\frac{0.960}{2} + 0.740 + 0.180 \right] = 0.233 \text{ ampere}$$

$$(\hat{I}_p)_1 = \frac{1}{6} [0.960 + 1.73 \cdot 0.740 + 0.180] = 0.404 \text{ ampere}$$

$$\bar{I}_c = \frac{1}{6} \left[\frac{0.110}{2} + 0.025 + 0.001 \right] = 0.014 \text{ ampere}$$

$$(\hat{I}_g)_1 = \frac{1}{6} [0.110 + 1.73 \cdot 0.025 + 0.0103] = 0.026 \text{ ampere}$$

Comparing the values of each component of current obtained by the long and short formulas shows that the agreement is close.

Using the results just obtained, the operating conditions for paths Qp_1 and Qp_2 as calculated from (7.7) through (7.14) are as listed in Table 11.1.

TABLE 11.1

Quantity	Path Qp_1	Path Qp_2
\bar{I}_b	0.236 ampere	0.552 ampere
\bar{I}_c	0.014 ampere	0.075 ampere
$(\bar{I}_p)_1$	0.402 ampere	0.920 ampere
$(\bar{I}_p)_1$	0.026 ampere	0.148 ampere
P_{bb}	$1,000 \cdot 0.236 = 236$ watts	552 watts
P_L	$\frac{800 \cdot 0.402}{2} = 161$ watts	367 watts
P_p	$236 - 161 = 75$ watts	185 watts
η_p	$\frac{161}{236} = 68.2$ per cent	66.5 per cent
P_d	$\frac{300 \cdot 0.026}{2} = 3.9$ watts	33.3 watts
P_{cc}	$150 \cdot 0.014 = 2.1$ watts	11.3 watts
P_g	$3.9 - 2.1 = 1.8$ watts	22.0 watts
A_p	$\frac{161}{3.9} = 41$	11

12. Selection of Path and Contour Diagrams for Fixed Q Point.

If the Q point is fixed but the end point p of the path is moved, the various factors that express the condition of operation of the tube vary. For example, moving p in Fig. 10.3 from p_1 to p_2 causes the changes shown in Table 11.1. For path Qp_1 the plate dissipation is 75 watts; this is less than the maximum safe dissipation of 100 watts for the particular tube, indicating that more power could be converted by this tube. For path Qp_2 , however, the plate dissipation of 185 watts far exceeds the allowable value. Furthermore, the grid driving power P_d is greater by a factor of nearly 10 and the grid dissipation is greater by a factor of over 12 when the path is moved from Qp_1 to Qp_2 . It is clear that, considering the heating of the tube elements, a path between the two would be safe.

Another factor may enter to limit the power input to the tube. As is mentioned in Sec. 2, the life of a thoriated-tungsten cathode may decrease as \bar{I}_b increases. For this reason the manufacturer usually places an upper limit to \bar{I}_b . For the 211 tube used for

illustration, this maximum \bar{I}_b is 175 ma. It appears therefore that even path Qp_1 is too far to the right.

In order to find a suitable path of operation it would be necessary to know by calculation or by measurement the various factors of operation for a large number of paths terminating at various points. The results then could be plotted so that a choice, determined by a consideration of all factors, could be made. The most convenient and lucid way of presenting and correlating the results is to show them in the form of contour diagrams.

The contour diagram for P_L now will be described. With a fixed position of the Q point, the end point p of the path of operation can be moved so that for each path the power output P_L is the same.

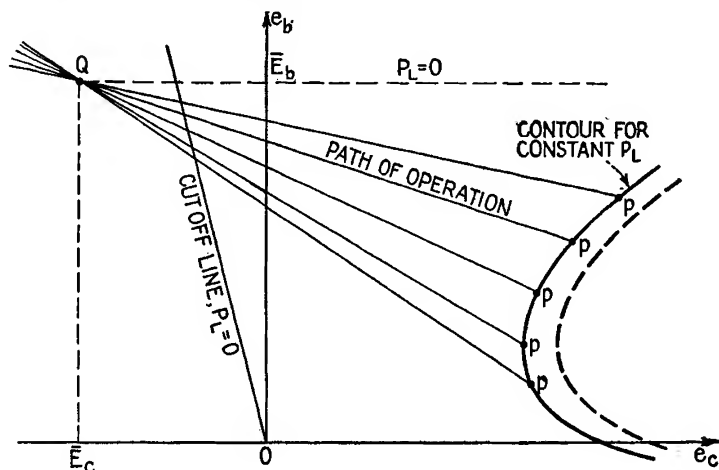


FIG. 12.1.—Contour for constant value of power output, P_L .

The locus of the end point, illustrated in Fig. 12.1, is the contour for this constant value of P_L . Contours for other values of P_L can be drawn. All will be of the same general shape as the one shown in Fig. 12.1 but shifted horizontally to the right for larger values of P_L and to the left for smaller values of P_L . The dashed-line contour in Fig. 12.1 is for a larger value of P_L than that for the solid-line contour. The cutoff line and the horizontal line through Q are the limits of these contours and together form the contour for P_L equal to zero, provided that Q is to the left of the cutoff. A family of contours for equal increments of P_L is a convenient way of showing the power output for any path of operation. Such a family of contours is shown in Fig. 12.3 for the particular tube whose static curves are given in Figs. 9.1 and 10.3.

Instead of moving the end point of the path of operation so as to maintain P_L constant, it may be moved in such a way as to maintain the plate dissipation constant. A contour of constant P_p

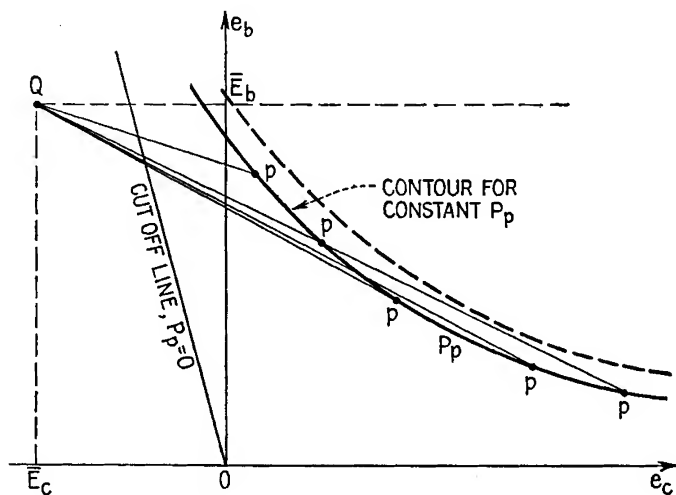


FIG. 12.2.—Contour for constant value of plate dissipation, P_p .

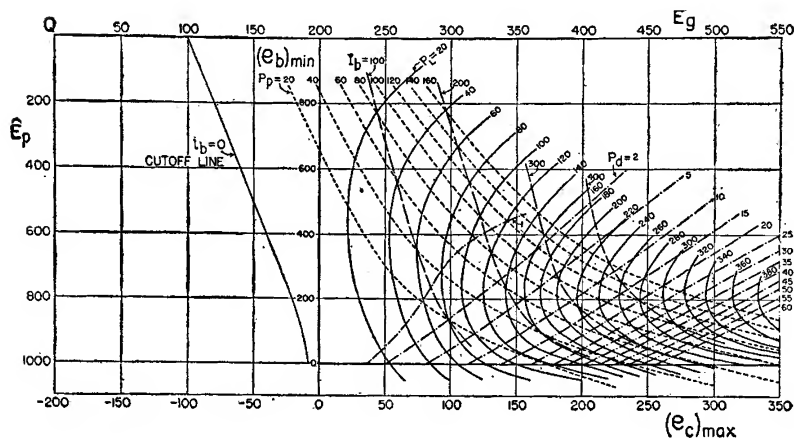


FIG. 12.3.—Chart showing contours of constant P_L , P_p , P_d , and I_b for the type 211 tube, Class C operation; $\bar{E}_c = -200$ volts.

is shown in Fig. 12.2. Contours for other values of P_p are approximately parallel to the one shown, as indicated by the dashed contour for a larger value of P_p . The contour for $P_p = 0$ coincides with the cutoff line. A family of contours for various values of

P_p is shown in Fig. 12.3. In a similar way, contours can be drawn for driving power P_d and for average plate current \bar{I}_b , Fig. 12.3.

A number of other contours can be drawn, some of which are shown in Fig. 12.4. For example, contours of constant plate-circuit efficiency η_p are lines approximately parallel to the e_c axis. Contours of P_g are of the same general shape as for P_d .

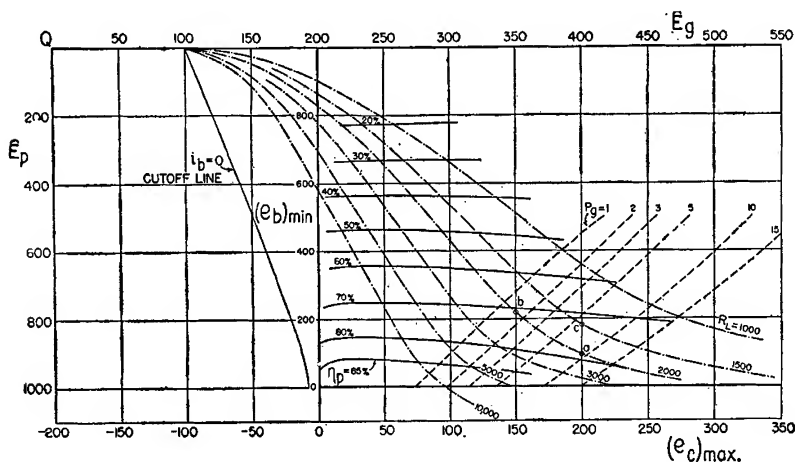


FIG. 12.4.—Chart showing contours of constant P_g , η_p , and R_L for the type 211 tube, Class C operation; $\bar{E}_c = -200$ volts.

Important contours are those for load resistance R_L , which may be calculated directly from contours of P_L , Fig. 12.3, by (7.8) rewritten as

$$R_L = \frac{\hat{E}_p^2}{2P_L} = \frac{\hat{E}_p}{(\hat{I}_p)_1} \quad (12.1)$$

These contours converge to the cutoff value of \bar{E}_c at the plate voltage \bar{E}_b . The R_L contour for the load used gives the locus of the end point of the path, and hence the value of \hat{E}_p as \hat{E}_g is varied.

The contours of Figs. 12.3 and 12.4 furnish a clear picture of the way in which the various powers and plate efficiency vary as the end point of the path is made to move in various directions. For example, if \hat{E}_g is decreased from 400 to 350 volts and the load resistance is 2,000 ohms, the end point moves from *a* to *b*, Fig. 12.4, the output P_L decreases from 200 to 150 watts, the value of η_p decreases from 78 to 71 per cent, P_d decreases from 19 to 4 watts, and P_p increases from 47 to 60 watts. On the other hand, if the driving voltage remains constant at the value of 400 volts and R_L is decreased from 2,000 to 1,500 ohms, the end point moves from

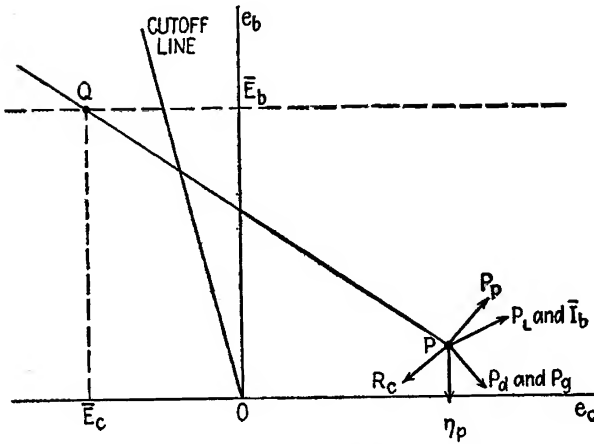


FIG. 12.5.—Direction of shift of the point p to give the most rapid increase of the various quantities indicated.

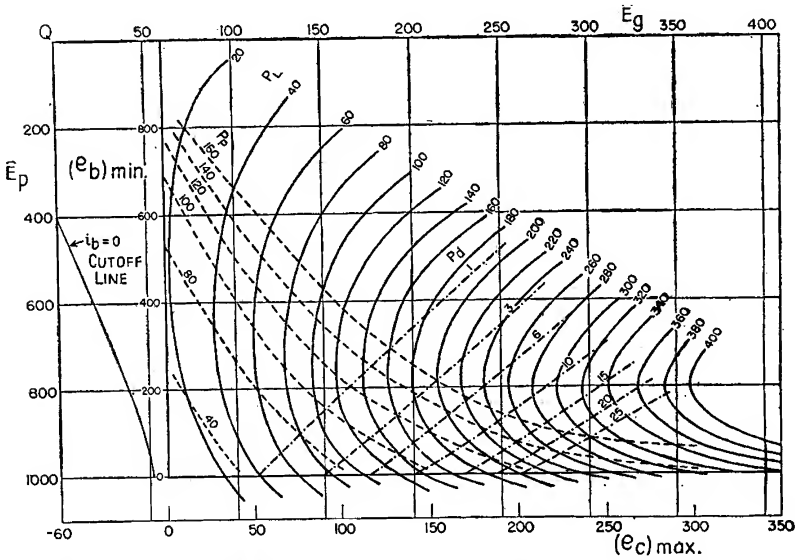


FIG. 12.6.—Chart showing contours of constant P_L , P_p , and P_d for the type 211 tube, Class B operation; $\bar{E}_c = -60$ volts.

a to c , Fig. 12.4, and the various powers change as shown by the contours.

The directions in which the end point of the path of operation must move to cause the most rapid increase in the various quantities are shown in Fig. 12.5. The directions depend upon the position of

the end point; for example, the direction of most rapid increase in P_L depends upon whether the end point lies above, at, or below the point on the P_L contour having a vertical tangent. The position of the end point chosen in Fig. 12.5 is in the region of most practical operation, which is nearly always below the level of the point on the P_L contour having a vertical tangent.

With a fixed Q point, the only way to move the position of the end point of the path is to change \bar{E}_g , or R_L , or both.

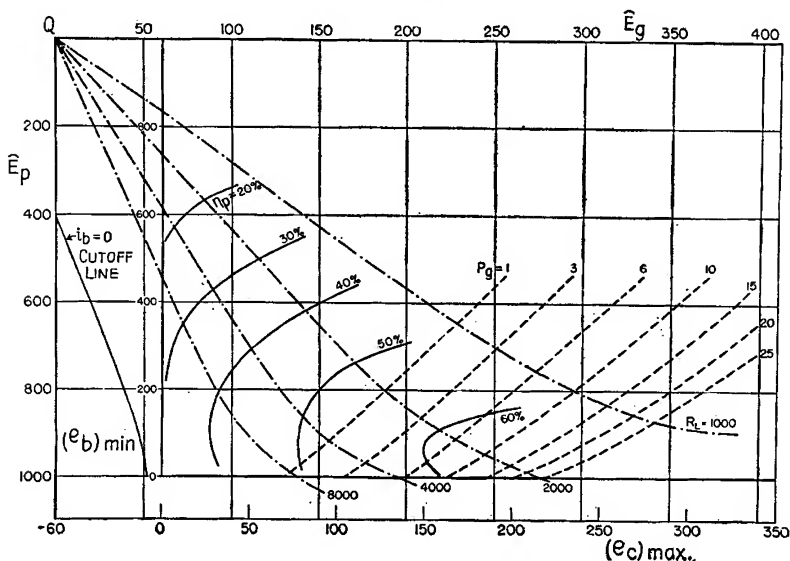


FIG. 12.7.—Chart showing contours of constant P_g , η_p , and R_L for the type 211 tube, Class B operation; $\bar{E}_c = -60$ volts.

If the position of the Q point is changed, all contours shift in position. For example, if the negative grid bias is decreased, the contours for P_L shift to the left and slightly downward, while the contours of P_p shift diagonally downward to the left. The contours for P_d shift diagonally to the lower right, and those for η_p shift downward. The amount of shift is generally proportional to the change in bias. Contours for a bias of -60 volts are shown in Figs. 12.6 and 12.7.

13. Neutralization.—If a triode is used as a power amplifier in a circuit as shown in Fig. 7.1 or 7.2, some high-frequency power passes from the grid circuit to the plate circuit, or vice versa, through the capacitance inside the tube between plate and grid electrodes and the capacitance outside the tube between the circuit

elements and wires connected to grid and plate. Such power fed back is generally positive and causes the stage to oscillate and hence to be unstable. To eliminate this power feedback, triodes must be "neutralized." Tetrodes and pentodes usually do not require neutralization because the screen grid reduces the grid-to-plate capacitance sufficiently to eliminate practically all this interchange of power between input and output circuits.

The principle of neutralization is very simple. There are a number of methods for its accomplishment in power-tube amplifiers.

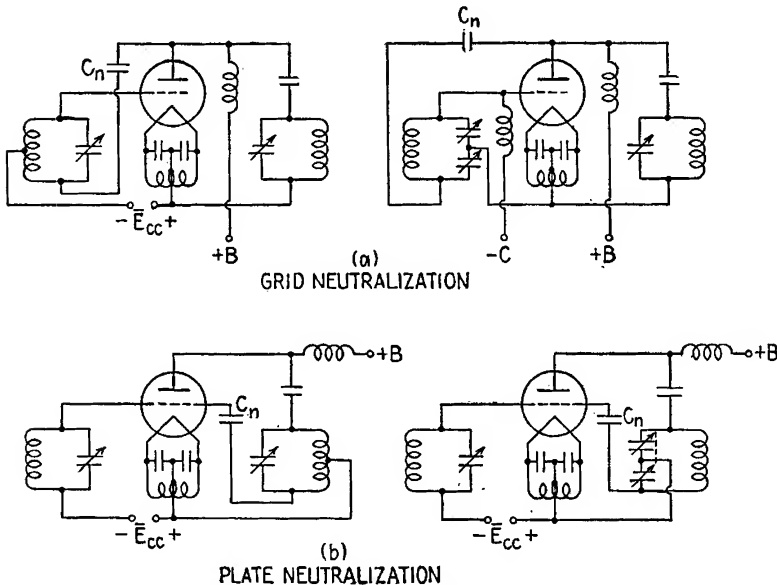


FIG. 13.1.—Neutralization of a single-ended amplifier stage.

To neutralize an amplifier stage, power must be fed by external circuits between the output and input circuits in amount equal to but opposite in phase to that which passes through C_{pg} .

In Fig. 13.1 are shown two commonly used connections for neutralization of a power-amplifier stage. In Fig. 13.1a the plate is connected through the adjustable neutralizing capacitance C_n to a point in the grid circuit that has a potential of opposite phase to that of the grid.¹ This method or connection is often called "grid neutralization." The scheme² shown in Fig. 13.1b,

¹ C. W. RICE, U.S. patent 1,334,118, filed July, 1917.

² L. A. HAZELTINE, U.S. patent 1,450,080, filed August, 1919; U.S. patent 1,489,228, filed December, 1920.

known as "plate neutralization," utilizes a point in the plate circuit having a potential of opposite phase to that of the plate. Figure 13.2 shows the simple method of neutralizing a push-pull stage.

The connections for neutralization just described are effective over a fair range of frequencies. A scheme that provides nearly

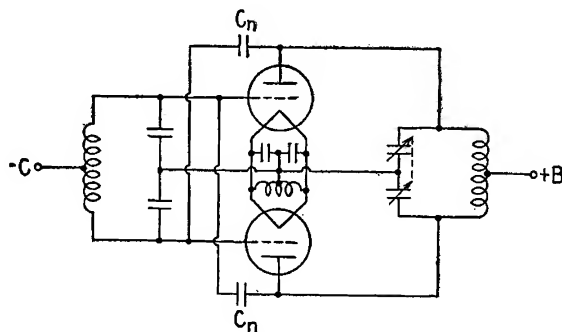


FIG. 13.2.—Neutralization of a push-pull amplifier stage.

complete neutralization for a single frequency is shown in Fig. 13.3. Capacitance C_n serves as a blocking capacitance. It also serves to tune the circuit $C_n L_n$ in parallel with C_{op} to parallel resonance, thus offering a high impedance between grid and plate electrodes. The advantage of this method is its simplicity and cheapness.

There are several methods of setting the neutralizing capacitance

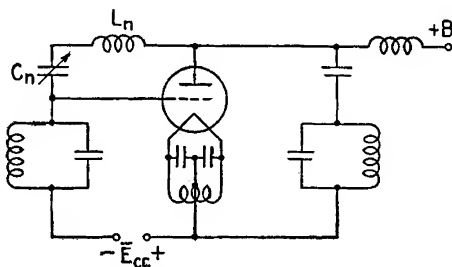


FIG. 13.3.—Simple method of neutralization at a single frequency.

C_n for best balance. The first method described may be applied when the cathode is either cold or hot although it is most sensitive in the former condition because the a-c grid voltage is then greater. For this test the plate-supply voltage should be turned off. The capacitance C_n is set so that no appreciable voltage is developed across the plate tank circuit as it is tuned to resonance with the

normal driving voltage applied to the grid circuit. Some sensitive device is used to detect the r-f voltage across the plate tank circuit, such as a cathode-ray oscilloscope connected across the tank circuit, a neon tube brought near to the plate terminal, a sensitive radio-frequency ammeter, or a small flashlight bulb included in a few turns of wire coupled to the tank-circuit inductance.

A second method of neutralizing, applicable when the cathode is hot, is to set the neutralizing capacitor so that tuning the plate tank circuit through resonance gives a minimum reaction on the average grid current \bar{I}_c , as indicated by a grid-current milliammeter. For this test the source of plate current should be disconnected in order that there be no electronic coupling between the circuits.

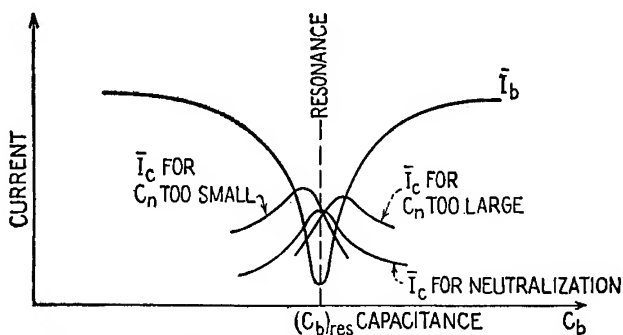


FIG. 13.4.—Effect of varying C_b upon the grid current for different values of C_n .

A third and very satisfactory method of neutralizing an amplifier is applied while the tube is functioning at full power and has the advantage therefore of balancing the capacitance C_{op} as it is affected by the normal electronic space charge. It was explained in Sec. 6 that detuning the plate tank capacitance in either direction causes an increase in average plate current and a decrease in average grid current, as shown in Fig. 6.7. If, however, the stage is not neutralized, the maximum grid current and minimum plate current do not occur at the same setting of tank capacitance. This effect is shown in Fig. 13.4. For proper neutralization, C_n is set so that both maximum grid current and minimum plate current occur at the same setting of the tank-circuit capacitance. To test for neutralization after adjusting C_n , C_b is altered slightly in both directions from the resonance value, but in making these alterations in C_b the change from the resonance value should be small to prevent overheating of the plate.

circuit increases and is of the order of 10 per cent or more of the fundamental voltage.

The tuning procedure for the circuit of Fig. 14.1 is as follows: With reduced plate-supply voltage, the capacitor C_c is adjusted to maximize the grid voltage \hat{E}_g as indicated by *maximum* grid current \bar{I}_c ; then the plate tank-circuit capacitor C_b is tuned for *minimum* plate current \bar{I}_b . If the tube is not properly loaded when the plate-supply voltage is increased to the normal value, the constants of the tank circuit should be changed as explained previously. The proper load is that which provides satisfactory

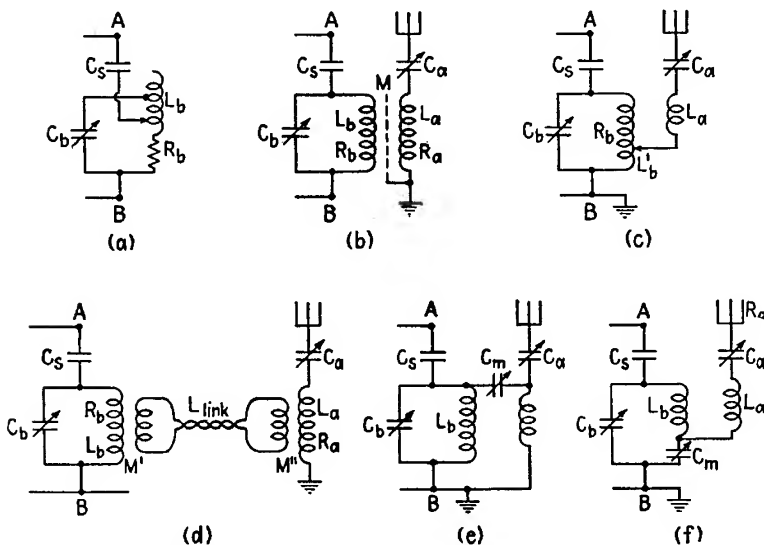


FIG. 14.2.—Load circuits for a power amplifier.

P_L without overheating the plate. The value of the plate current \bar{I}_b is the indication of the loading of the tube when and only when the load has unity power factor.

Instead of changing the constants of the tank circuit in order to obtain the proper R_L , the impedance of the load between A and B can be varied by changing the position of the plate connection to the tank circuit, Fig. 14.2a. For every position of the tap, C_b must be reset for a minimum \bar{I}_b . This method of varying R_L has the one objectionable feature, that the harmonic components of plate current do not have a capacitive path to cathode as they do in the circuit of Fig. 14.1 and hence the plate voltage may contain excessive harmonic components.

If the power is to be fed to a second circuit, for example to an antenna, the second circuit may be inductively coupled to the tank or plate circuit, Figs. 14.2*b*, *c*, *d*, or capacitively coupled, Figs. 14.2*e*, *f*. The direct coupling of Fig. 14.2*c* is merely a special form of the scheme of Fig. 14.2*b*. The link coupling shown in Fig. 14.2*d* is equivalent to the connection of Fig. 14.2*b* but enables the antenna and tank to be separated by a short distance. Either of the mutual inductances M' or M'' in the link circuit may be provided by direct coupling instead of by mutual coupling. The connections of Figs. 14.2*e*, *f* are equivalent except that the latter provides better harmonic suppression. Other more complicated coupling connections are sometimes used to obtain increased harmonic suppression, but the fundamental principles of operation are the same as for the simpler circuits.

In all the coupled output circuits the condition for maximum efficiency of power transfer from the tank circuit to the coupled system is that the *circuit coupled to the tank be tuned to resonance*.¹ Then the coupling interposes *only a resistance* into the tank circuit.

Considering first the magnetically coupled circuit, Fig. 14.2*b*, the required condition of antenna tuning is

$$X_a = L_a\omega - \frac{1}{C_a\omega} = 0 \quad (14.4)$$

The load resistance R_L is then

$$R_L = \frac{L_b}{C_b \left(R_b + \frac{\omega^2 M^2}{R_a} \right)} \quad (14.5)$$

or

$$R_L = \frac{L_b}{C_b R_b} \frac{1}{1 + (k/k_e)^2} \quad (14.6)$$

In the circuit of Fig. 14.2*c* the condition of maximum efficiency of power transfer demands that the entire antenna circuit be tuned to resonance, or

$$\omega(L_a + L_b') - \frac{1}{\omega C_a} = 0 \quad (14.7)$$

In the link-coupled circuit of Fig. 14.2*d*, the circuit coupled to the tank circuit should be considered as an equivalent circuit comprising the link circuit into which is reflected the antenna

¹ Chap. VII, Sec. 21.

impedance. The condition that the circuit be tuned to resonance is

$$\omega L_{\text{link}} - \frac{\omega^2 M'^2 X_a}{R_a^2 + X_a^2} = 0 \quad (14.8)$$

This relation shows that the reactance X_a of the antenna circuit cannot be zero but must be sufficiently positive so that the reactance coupled into the link circuit is sufficiently negative to cancel the positive reactance of the link circuit. The load resistance is then

$$R_L = \frac{L_b}{C_b \left(R_b + \frac{\omega^2 M'^2}{R_{\text{link}} + \frac{\omega^2 M'^2 R_a}{R_a^2 + X_a^2}} \right)} \quad (14.9)$$

For the capacitively coupled circuit of Fig. 14.2*f*, the condition of tuning for maximum circuit efficiency is

$$X_a = \omega L_a - \left(\frac{1}{\omega C_a} + \frac{1}{\omega C_m} \right) = 0 \quad (14.10)$$

and the load resistance R_L will be

$$R_L = \frac{L_b}{C_b R_b} \frac{1 - \frac{1}{L_b C_m \omega^2}}{1 + \frac{1}{R_b R_a C_m^2 \omega^2}} \quad (14.11)$$

$$= \frac{L_b}{C_b R_b} \frac{1 - \frac{1}{L_b C_m \omega^2}}{1 + (k/k_c)^2} \quad (14.12)$$

In (14.6) and (14.12), k_c represents critical coupling for the circuits when the cathode of the tube is cold, and k is the coupling coefficient used; the coupling lowers R_L by dividing the value of R_L of the tank circuit alone by the factor $1 + (k/k_c)^2$. It is shown in Chap. VII, Sec. 21, that for good efficiency of the circuits this factor should be not much less than 26, corresponding to a value of 5 for k/k_c .

The tuning procedure for the coupled output circuits is somewhat more involved than for the singly tuned tank circuit. With reduced plate voltage the first step is to set the grid capacitor C_c for a *maximum* \bar{I}_c ; then, with the coupled circuit open or decoupled, the tank capacitor C_b is tuned for a *minimum* \bar{I}_b . The coupled circuit then is closed or coupled and is tuned for a *maximum antenna* current. This latter tuning may have detuned the load away from unity power factor. This will occur in circuits such as are

shown in Figs. 14.2*e, f*, where changing the coupling element C_m also affects the tuning of the tank circuit. To correct the power factor of the load, C_b now should be reset for a *minimum* \bar{I}_b . The plate voltage now may be increased to the normal value, the tube or plate milliammeter being watched to ensure that the tube is not overloaded. If the loading is not correct, the coupling should be changed and the tuning procedure repeated. It is well to emphasize again that the setting of C_b for a minimum \bar{I}_b should be the *final* adjustment, as this condition ensures that the phase angle between E_p and E_o is 180° , hence that the tube is operating at its maximum efficiency for the particular value of load resistance.

The tuning procedure just described is simple and practical if a good antenna ammeter is available. Frequently, however, the reading on the antenna ammeter is too small to be effective as an indicator, or the ammeter is too sluggish, or it is not conveniently located. In the procedure described, when the antenna system is tuned for maximum antenna current the average plate current is also large. The rise in plate current occurs because the greatest resistance is reflected into the tank circuit when the coupled circuit is in resonance. The greater the reflected resistance, the lower the load impedance and the lower the opposition to the flow of plate current. It would seem possible then to use the maximum plate current as an indication of resonance of the coupled circuit, rather than the maximum antenna current. Sometimes this is possible, but often maximum plate current does not occur simultaneously with maximum antenna current. This is because the magnitude of the plate current depends not only upon the magnitude of the load impedance but also to a considerable extent upon the power factor of the load. The complication that arises in using the plate-current meter as the sole indicator of coupled circuit tuning can best be understood from the accompanying diagrams. In Fig. 14.3, the axes are the values of the tuning elements, C_b in the tank circuit and C_a in the coupled circuit or antenna. The specific value of each capacitance for resonance of each circuit by itself is indicated. If C_a is given any value and C_b is set for a minimum \bar{I}_b , the curve marked "Locus of min. \bar{I}_b " is obtained. This curve corresponds closely to the curve for partial resonance with secondary fixed, Chap. VII, Sec. 10. It is the curve for unity power factor for the plate load, and its shape depends upon k and Q_a . The other curve gives the value of C_a , for any preset value of C_b , to give maximum \bar{I}_b .

The tuning procedure, assuming k to be fixed and starting from any settings of C_a and C_b , as for example at A , is as follows: First C_b is tuned for a minimum \bar{I}_b at point a ; next C_a is set to give a maximum \bar{I}_b at b . Capacitances C_b and C_a are alternately readjusted to points c, d, e, f , etc., until the proper tuning point is reached at the intersection of the two curves.

When the coupling is increased or Q_a is increased, the two loci assume the relative positions shown in Fig. 14.4, where angle θ_{\max} is greater than angle θ_{\min} . In this condition the process of tuning, starting from some point such as A , leads to a divergent

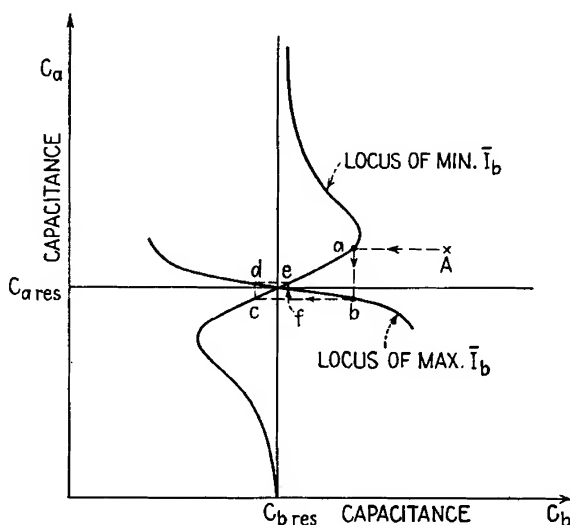


FIG. 14.3.—Convergent tuning using \bar{I}_b as indicator.

condition, as shown by the figure. Such a procedure does not lead in this case to the correct settings of the circuits, but it leads to two alternative conditions neither of which is a correct tuning condition.

When the divergent tuning condition prevails, the correct tuning condition can be reached if, after each adjustment to a condition represented by the loci, the setting of each capacitance is backed off slightly toward its former setting. This method of tuning is shown graphically in Fig. 14.5, by the sequence denoted by A, a, b, c, d, e , etc. This procedure is called the *back-off method*. An alternative procedure, called the *incremental method*, is shown by the sequence A, a, b', c', d', e' , etc. In this procedure the antenna capacitance is changed only slightly in the direction to increase \bar{I}_b ,

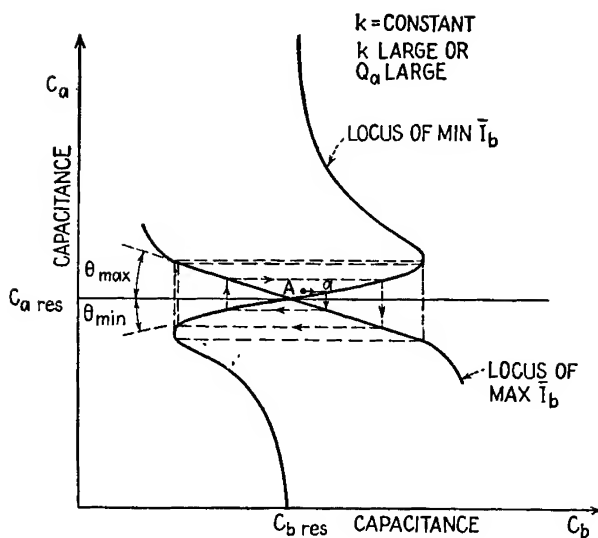
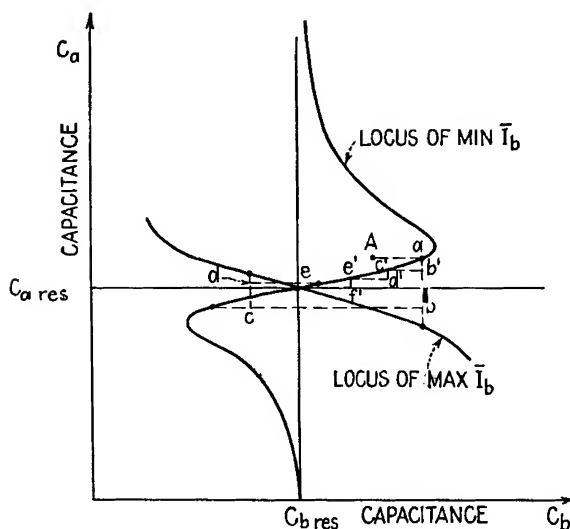
FIG. 14.4.—Divergent tuning using \bar{I}_b as indicator.

FIG. 14.5.—Procedures for tuning coupled systems (a) back-off method, path A, a, b, c, d, etc.; (b) incremental method, path A, a, b', c', d', etc.

and then the tank capacitance is reset for a minimum \bar{I}_b . This cycle is repeated until no increase in \bar{I}_b is obtained when the antenna capacitance is changed slightly. This second procedure is preferred, for it does not cause the excessive heating of the plate that may result in the back-off method.

Frequently the power is fed to the antenna through a nonresonant transmission line. Either coaxial or parallel-wire lines may be used, and various methods of coupling to the line and to the antenna are possible. Figure 14.6 shows two typical connections, the first using a coaxial line and the second a higher impedance two-wire line.

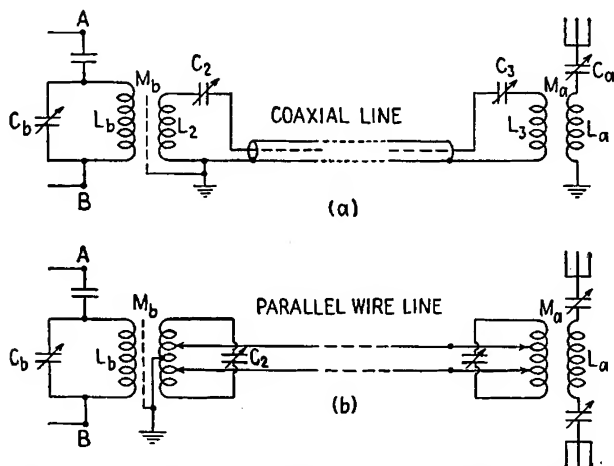


FIG. 14.6.—Typical nonresonant connections to an antenna.

The tuning procedure of systems employing nonresonant lines breaks down into two independent parts, *viz.*, tuning the antenna end and tuning the tube end. The tuning of the antenna end of the system depends upon the characteristic resistance of the line and must be done so as to terminate the line by this characteristic resistance. This impedance match may be accomplished by T or Π networks, Chap. V, by magnetically coupled circuits, Chap. VII, or by line sections or stubs.¹ The tuning of the tube end of the system is the same as it would be if the line were replaced by a resistance equal to the characteristic resistance of the line.

¹ R. W. P. KING, H. R. MIMNO, and A. H. WING, "Transmission Lines, Antennas, and Wave Guides," Chaps. I, II, McGraw-Hill Book Company, Inc., 1945.

The correct tuning of the system of Fig. 14.6a results in the following conditions:

$$X_a = \omega L_a - \frac{1}{\omega C_a} = 0 \quad \left. \begin{array}{l} \text{for maximum efficiency} \\ \text{of power transfer} \end{array} \right\} \quad (14.13)$$

$$X_3 = \omega L_3 - \frac{1}{\omega C_3} = 0 \quad \left. \begin{array}{l} \text{to terminate the line} \\ \text{in a resistance} \end{array} \right\} \quad (14.14)$$

$$R_c = R_3 + \frac{\omega^2 M_a^2}{R_a} \quad \left. \begin{array}{l} \text{for proper line termination} \end{array} \right\} \begin{array}{l} \text{Antenna} \\ \text{end} \end{array} \quad (14.15)$$

$$X_2 = \omega L_2 - \frac{1}{\omega C_2} = 0 \quad \left. \begin{array}{l} \text{for maximum efficiency} \\ \text{of power transfer} \end{array} \right\} \quad (14.16)$$

$$\frac{L_b}{C_b} = \left(R_b + \frac{\omega^2 M_a^2}{R_2 + R_c} \right)^2 + \omega^2 L_b^2 \quad \left. \begin{array}{l} \text{for unity} \\ \text{power factor} \end{array} \right\} \quad (14.17)$$

$$R_L = \frac{L_b}{C_b \left(R_b + \frac{\omega^2 M_a^2}{R_2 + R_c} \right)} \quad \left. \begin{array}{l} \text{for proper tube loading} \end{array} \right\} \begin{array}{l} \text{Tube} \\ \text{end} \end{array} \quad (14.18)$$

15. Calculation of Output Circuits of a Power Amplifier.—The load circuits of a power amplifier are designed to provide the specified value of R_L as determined by the contour charts or by calculation and to tune to a given value of ω . These two conditions are not sufficient to determine the values of the equivalent L_b , C_b , R_b of the tank circuit. Another condition must be added. This third condition may be a fixed value of one of the elements such as R_b chosen because of some external considerations. Generally, such fixed conditions may be embodied in assigning a value to the equivalent Q of the tank circuit. In order that the ratio of harmonic voltage to fundamental voltage be not excessive, the equivalent Q of the load should be not less than 10, say. The design relations can be derived from the following fundamental equations for the circuit of Fig. 14.1. Upon combining (14.2) and (14.3) with $Q_b = \omega L_b / R_b$ at the parallel-resonant frequency, these equations give the following relations,

$$\omega C_b = \frac{Q_b}{R_L} \quad (15.1)$$

$$\omega L_b = \frac{Q_b R_L}{1 + Q_b^2} \quad (15.2)$$

$$R_b = \frac{R_L}{1 + Q_b^2} \quad (15.3)$$

The value of R_b in these relations is the effective resistance of the tank circuit and is greater than the resistance of the coil alone by the reflected resistance from the coupled antenna circuits. In place of R_b the quantity

$$R_b + \frac{\omega^2 M^2}{R_a} = R_b \left[1 + \left(\frac{k}{k_c} \right)^2 \right] \quad (15.4)$$

must be used for the magnetically coupled circuits, and corresponding values for capacitively coupled systems. The value of k/k_c should be such as to give a reasonably high value of the circuit efficiency, as given by (21.7) of Chap. VII,

$$\eta = \frac{(k/k_c)^2}{1 + (k/k_c)^2} \quad (15.5)$$

A satisfactory value for $1 + (k/k_c)^2$ is 26.

As an illustrative example, constants of the circuits of Fig. 14.2b will be calculated for the conditions for path Qp_1 of Fig. 10.3 given in the first column of Table 11.1. Assuming Q_b to be 10 and ω to be 10^7 radians/sec and taking from Fig. 10.3 that $\hat{E}_p = 800$ volts, R_L is, from (12.1),

$$R_L = \frac{800^2}{2 \cdot 161} = 1,990 \text{ ohms} \quad (15.6)$$

From (15.1),

$$C_b = \frac{10}{1,990 \cdot 10^7} = 503 \text{ } \mu\text{f} \quad (15.7)$$

From (15.2),

$$L_b = \frac{10 \cdot 1,990}{(1 + 100) \cdot 10^7} = 19.7 \text{ } \mu\text{h} \quad (15.8)$$

From (15.3) and (15.4),

$$R_b \left[1 + \left(\frac{k}{k_c} \right)^2 \right] = \frac{1,990}{1 + 100} \quad (15.9)$$

Assume that the tank coil alone¹ has a Q_b of 400. Then

$$R_b = \frac{L_b \omega}{Q_b} = \frac{197}{400} = 0.49 \text{ ohm} \quad (15.10)$$

¹ Data on coils is given in F. E. TERMAN, "Radio Engineers' Handbook," p. 74, McGraw-Hill Book Company, Inc., 1943.

Substituting the value of R_b in (15.9),

$$\left(\frac{k}{k_c}\right)^2 = \frac{1,990}{(1 + 100) \cdot 0.49} - 1 = 39.0$$

or

$$k = 6.24k_c$$

According to (15.5), the efficiency of the circuits when properly tuned is $\frac{39}{40}$, or 97.5 per cent. Assume that the antenna resistance R_a is 20 ohms. Then, since $\omega^2 M^2 / (R_a R_b) = (k/k_c)^2 = 39.0$,

$$M = \frac{1}{10^7} \sqrt{39 \cdot 20 \cdot 0.49} = 1.95 \mu\text{h}$$

The design of the antenna circuit will not be completed here. It is necessary to have sufficient inductance in the antenna-coupling coil to obtain the necessary mutual inductance M .

16. Class C Amplifier.—The Class C amplifier is one for which the grid-bias voltage is so large that plate current flows for *less* than one-half of the cycle. The Q point for a Class C amplifier is located to the left of the cutoff line as in Fig. 10.3. The greater the negative bias, the higher the plate-circuit efficiency but the greater the driving power. A good compromise dictates a bias voltage of 1.5 to 3 times the cutoff grid voltage.

Class C amplifiers are used for high-efficiency amplification of radio-frequency power having a constant amplitude or power that is interrupted as in sending a code message by continuous waves. Class C amplifiers are also used to modulate radio-frequency power and to amplify frequency-modulated power.

The determination of the conditions for satisfactory performance is explained in Sec. 12.

17. Class B Amplifier.—A Class C amplifier is unsuited to the amplification of a modulated signal, as is shown by Fig. 17.1. The grid bias is taken to be three times cutoff value. The resistance line for R_L is the locus of the end of the path of operation as the amplitude \hat{E}_g of the alternating component of the grid voltage varies. This line, therefore, gives the relation between \hat{E}_p and \hat{E}_g . The envelope of the modulated output voltage can be derived by projection from the envelope of the grid voltage to the R_L line as illustrated in the figure. Evidently a completely modulated input wave results in an overmodulated output wave.

In order to obtain an undistorted modulation of the output wave the resistance line must be a straight line passing through the

Q point. Since the resistance lines emanate from the cutoff point, the Q point must be moved to the right. The most nearly linear R_L line is obtained by using a grid bias somewhat *less* than cutoff. Figures 12.6 and 12.7 show the contours for the proper bias of an amplifier for a modulated voltage. Such an amplifier passes plate current for approximately one-half cycle and is called a Class B radio-frequency amplifier.

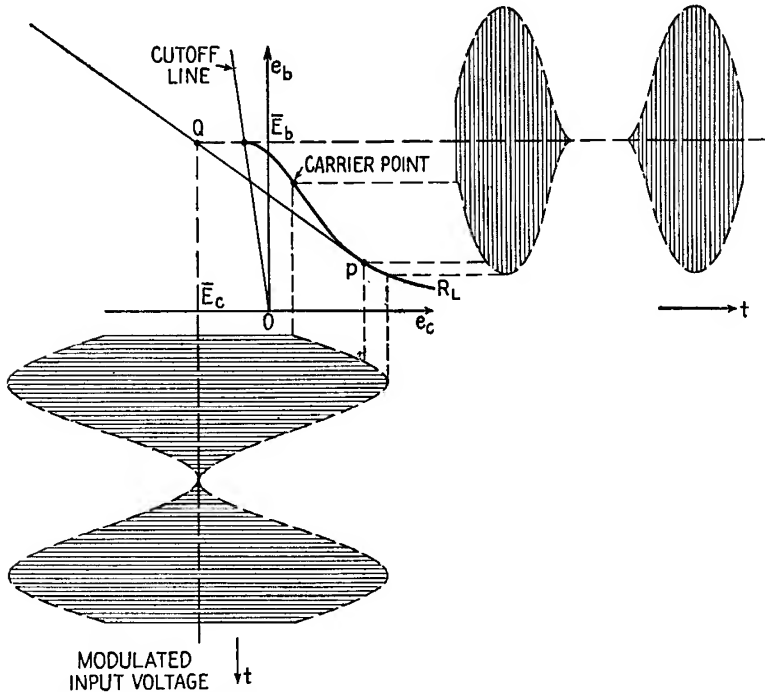


FIG. 17.1.—Class C amplifier used for a modulated signal, showing resulting distortion.

Examination of Fig. 12.7 shows that although the resistance lines are quite straight over their upper portion they all curve in their lower portion where the alternating grid and plate voltages are high. This bending cannot be avoided and results in distortion for high degrees of modulation.

The mode of operation of a Class B amplifier is shown in Fig. 17.2. The distortion at the peak values of output voltage is shown by the departure from the sinusoidal envelope that would result if the R_L line were straight throughout the operating range.

The carrier point, or the condition for no modulation, is indicated

by point C . At this point the plate-current efficiency is very low, of the order of 30 to 35 per cent as can be seen by examination of Fig. 12.7. The average efficiency at best of a Class B amplifier is consequently not over 40 to 45 per cent.

A high-efficiency Class B system of amplification has been invented by Doherty.¹

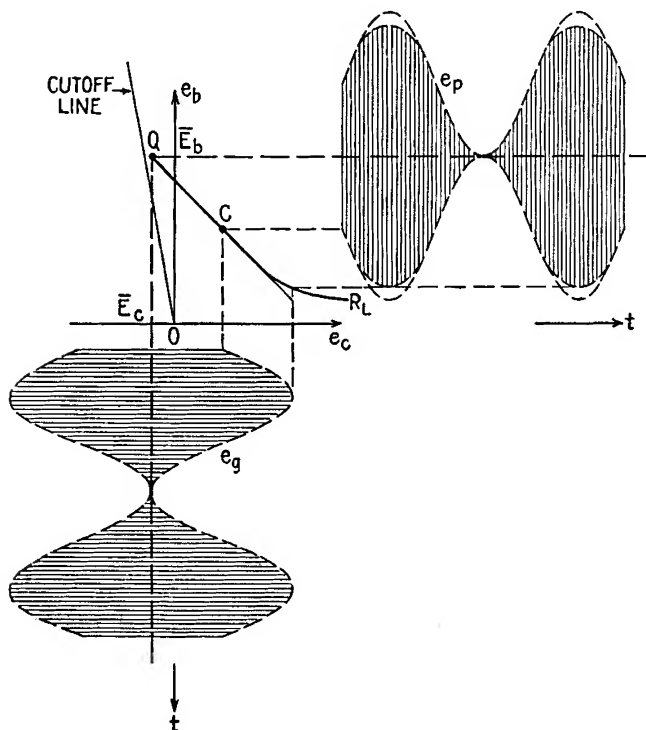


FIG. 17.2.—Class B amplifier.

18. Grid-modulated Power Amplifier.²—Various methods of modulation are described in Chap. XX. This section presents a further discussion of grid modulation of a power amplifier treated more from the point of view of the operation of the tube as it has been discussed in the immediately preceding sections.

A grid-modulated power amplifier should be thought of as a Class C power amplifier the grid bias of which is caused to vary

¹ W. H. DOHERTY, A New High-efficiency Power Amplifier for Modulated Waves, *Proc. I.R.E.*, **24**, 1163, 1936.

² See Chap. XIX, Secs. 1-4, and Chap. XX, Secs. 2-5 before reading Secs. 18 and 19 of this chapter.

in accordance with the modulation. This variation is considered to be slow as compared with the radio-frequency changes. A simple diagram of connection for the system is shown in Fig. 18.1. The only change necessary from the power-amplifier diagram of Fig. 14.1 is the addition of a modulating voltage connected in series with the grid-bias voltage \bar{E}_{cc} .

The mode of operation of this system is shown in Fig. 18.2. The steady part of the grid bias \bar{E}_c is given by the bias source \bar{E}_{cc} . The low-frequency variation in grid bias caused by the modulating

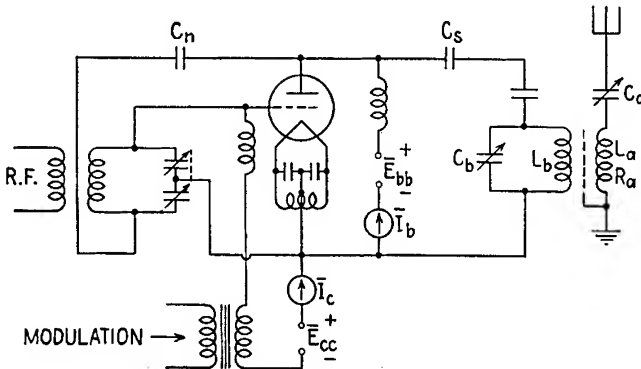


FIG. 18.1.—Connections for grid modulation of a Class C amplifier.

voltage is represented by a sine wave of amplitude $(\hat{E}_g)_l$. The r-f or high-frequency grid voltage is $(\hat{E}_g)_h$. The total instantaneous grid voltage is

$$e_c = \bar{E}_c + (\hat{E}_g)_h \sin \omega_h t + (\hat{E}_g)_l \sin \omega_l t \quad (18.1)$$

When the grid bias is $-\bar{E}_{cc}$, half the r-f path of operation is Q_0p_0 . When the grid bias is $-\bar{E}_{cc} + (\hat{E}_g)_l$ the half path is Q_1p_1 , and when the bias is $-\bar{E}_{cc} - (\hat{E}_g)_l$ the half path is Q_2p_2 . Each of these paths is for the same load resistance R_L , and hence each path must have the same value of $\hat{E}_p/(\hat{I}_p)_1$. The positions of the end points can be determined easily by means of the harmonic analysis and the static curves so that the above ratio is constant. For intermediate values of grid bias during the modulation cycle the end points p travel over a locus indicated in Fig. 18.2. Projection from this locus gives the envelope of the output voltage \hat{E}_p .

Since the locus of the end points just described is never straight up to the cutoff point or at the high-power end, distortion of the envelope is inevitable at high degrees of modulation.

The variations in the operation of the system can be predicted from the diagram of Fig. 18.2, where $(\bar{E}_g)_i$, or $(\bar{E}_g)_h$, or R_L is given various values. The power output, efficiency, driving power, plate dissipation, and other quantities can be calculated by the application of the harmonic analysis.

Grid modulation possesses the great advantage that little modulating power is required although this advantage is generally counterbalanced by the distortion always present, especially for high degrees of modulation.

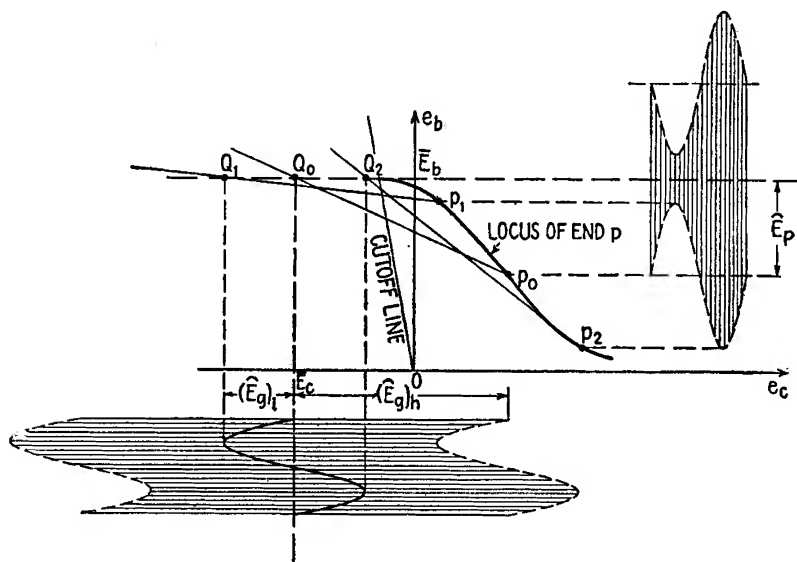


FIG. 18.2.—Grid modulation of a Class C amplifier.

19. Plate-modulated Power Amplifier.—A power amplifier also can be modulated by varying the plate-supply voltage instead of the grid-bias voltage, as described in Chap. XX, Sec. 2. Typical connections for this method of modulation are shown in Figs. 19.1a, b. In Fig. 19.1a the modulating voltage is introduced into the parallel-feed circuit by a transformer M . The primary coil of this transformer may be fed from any source such as a push-pull Class B audio amplifier. In Fig. 19.1b the modulation voltage is the voltage developed across the choke coil in the plate circuit of the modulator tube T_m . This latter connection is often known as the Heising constant-current system of modulation. This connection suffers from the disadvantage that the r-f amplifier cannot

be completely modulated unless the steady plate voltage of the r-f amplifier is reduced by a dropping resistance R or by some other equivalent scheme.

The mode of operation of a plate-modulated amplifier is shown

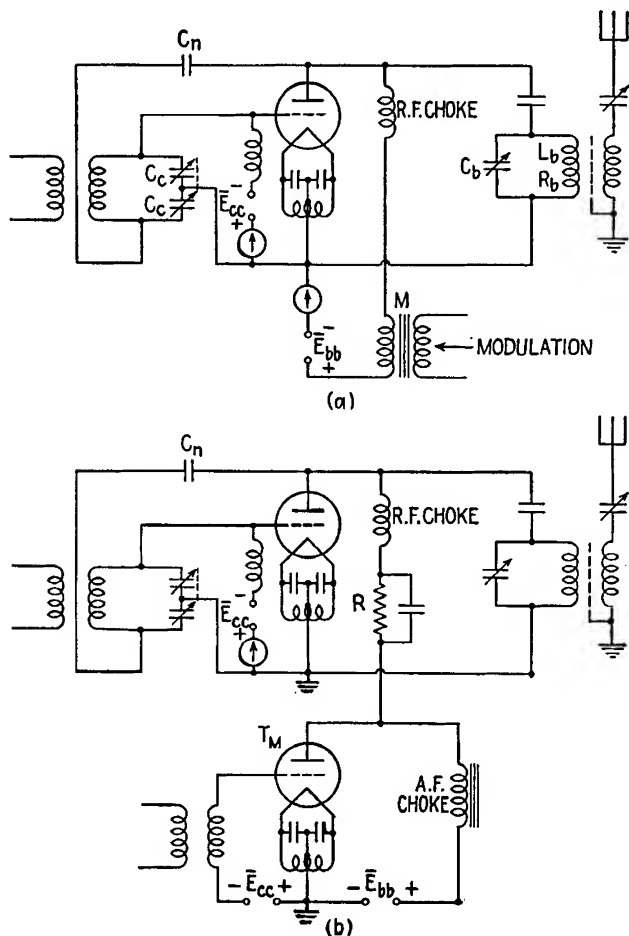


FIG. 19.1.—Connections for plate modulation of a Class C amplifier.

in Fig. 19.2. A constant r-f grid voltage $(\bar{E}_g)_h$ is impressed in the grid circuit in the usual way for an r-f amplifier. The plate voltage varies about a steady value \bar{E}_b by the amount of the modulating voltage, shown in Fig. 19.2 as a sinusoidal voltage of amplitude $(\bar{E}_p)_l$. If $(\bar{E}_p)_l$ equals \bar{E}_b , 100 per cent modulation is obtained. The path of operation takes on various positions as the Q points

oscillate up and down between points Q_1 and Q_2 . The two extreme half paths are shown as Q_1p_1 and Q_2p_2 , while the path for the unmodulated carrier is Q_0p_0 . As in grid modulation, each path is such that $\hat{E}_p/(\hat{I}_p)_1$ has the same value, which is equal to the plate-load resistance R_L . The plate voltage undergoes fluctuations both at radio frequency and at modulation frequency, but owing to resonance only the modulated r-f component appears across the tank circuit in the connections of Fig. 19.1.

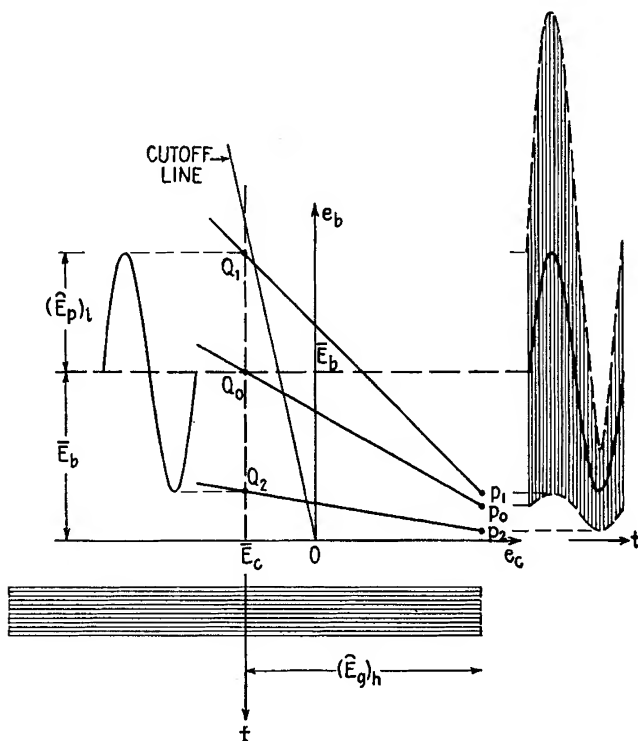


FIG. 19.2.—Plate-modulated Class *C* amplifier.

If the vertical spacings between the end points p , Fig. 19.2, are proportional to the spacing between the corresponding Q points, the modulation is undistorted. Generally this is not rigorously true as shown in Fig. 19.3, where the output voltage $(\hat{E}_p)_h$ is plotted against the voltage of the Q point, indicated by e_{bb} , for various values of R_L . The more nearly straight the lines, the less the modulation distortion. A high value of R_L gives less distortion but at the same time less power output.

Plate modulation is generally less distorting than grid modulation but requires much greater modulating power. For 100 per cent sinusoidal modulation the modulating power required is 50 per cent of the d-c input power from the plate-voltage supply (Chap. XX, Sec. 2).

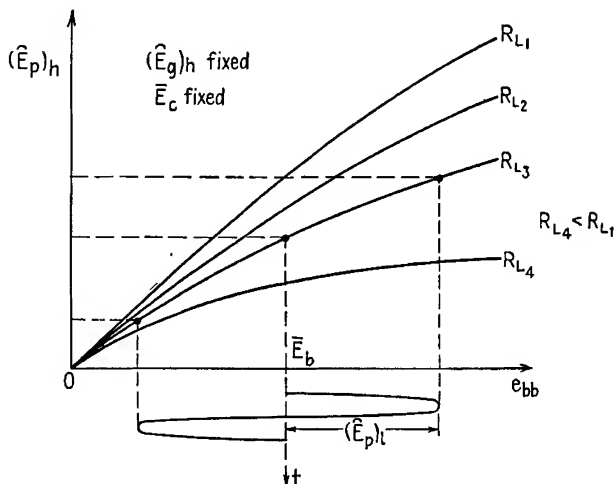


FIG. 19.3.—Modulation characteristics of a plate-modulated amplifier.

20. Oscillators.¹—Oscillators are discussed in Chap. XV. The purpose of this section is to add to the treatment there given certain considerations based upon the study of power amplifiers as developed in the present chapter. The discussion here will be concerned only with power oscillators.

Since a power oscillator is merely a self-fed power amplifier, the contour charts explained in Sec. 12 can be used for an oscillator provided that the driving power is subtracted from P_L to give the output power P_{LO} . For example, the contour chart of Fig. 20.1 may represent either an amplifier or an oscillator. Let it be assumed that point p is shown, as for an amplifier, to give satisfactory power P_L , plate dissipation P_p , and satisfactory values of other quantities such as η_p , P_d , etc. As an oscillator, the available power output P_{LO} is

$$P_{LO} = P_L - P_d \quad (20.1)$$

The value of P_p is that read off the contour diagrams. The grid bias is generally provided by the average grid current \bar{I}_c flowing

¹ See Chap. XV, Secs. 1–7, before reading Sec. 20 of this chapter.

through a biasing resistance R_c . The value of R_c is given by \bar{E}_c/\bar{I}_c . Since the grid driving voltage \hat{E}_g is derived from the plate output power P_L , which in turn depends upon \hat{E}_p , \hat{E}_g must bear a certain ratio to \hat{E}_p as determined by the circuits. The ratio \hat{E}_g/\hat{E}_p is the excitation ratio and fixes the slope of the path of operation as indicated in Fig. 20.1.

A resistance line can be drawn representing the equivalent resistance R_{LO} of the load circuit. This resistance differs from R_L for the amplifier because the whole of P_L of an amplifier goes into

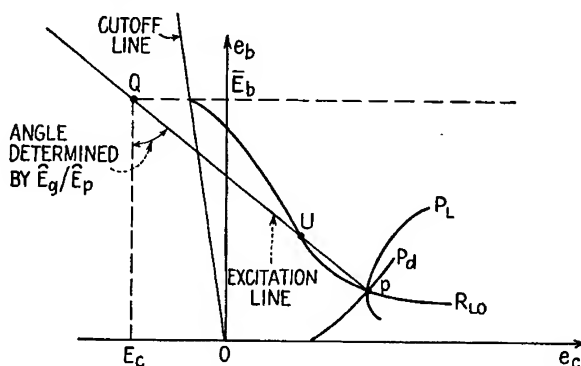


FIG. 20.1.—Contour chart showing conditions for an oscillator.

R_L , whereas in an oscillator the power $P_L - P_d$ is dissipated in R_{LO} . The following relations express these statements:

$$R_L = \frac{\hat{E}_p^2}{2P_L} \quad (20.2)$$

$$R_{LO} = \frac{\hat{E}_p^2}{2(P_L - P_d)} \quad (20.3)$$

A contour for R_{LO} can be drawn as in Fig. 20.1. Since the end of the path of operation must lie on this line, the point p giving the amplitude \hat{E}_p of oscillation is at the intersection of the excitation line and the R_{LO} line. The other point of intersection u represents an unstable condition. If the excitation line and the R_{LO} line do not intersect, steady oscillation is not possible. Any change in the circuits that decreases the excitation ratio, or increases the grid bias, or decreases the load resistance R_{LO} may make such an intersection impossible, resulting in a cessation of oscillation. When oscillation stops, the grid bias drifts back toward zero at a rate dependent upon the time constant of R_c and its shunting capacitance. Oscillation may start again as the bias diminishes, only to

stop again when the excitation line and R_{LO} cease to intersect. Such an action results in an intermittent, or "blocking," oscillator. The frequency of interruption may be low in the audio range or may be very high depending upon the time constant of the parallel combination of R_c and C_c . Intermittent operation is accentuated by increasing R_c , or by decreasing the excitation ratio, or by decreasing R_{LO} .

The path of operation of the oscillator shown in Fig. 20.1 is a straight line, implying that E_p and E_g are opposite in phase. Although this ideal phase relation may be secured by proper tuning in an amplifier, it does not follow that in an oscillator E_p and E_g

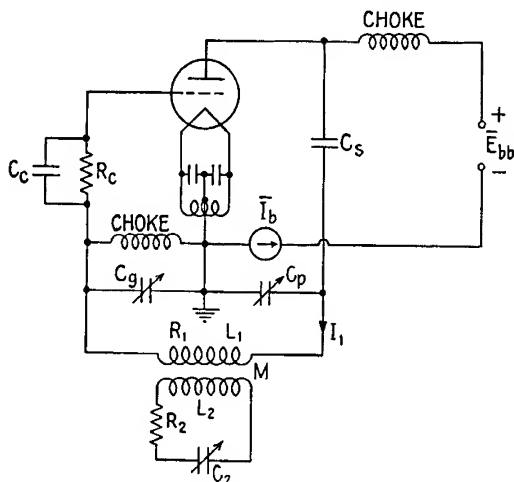


FIG. 20.2.—Colpitts oscillator.

are opposite in phase. For example, in a tuned-grid tuned-plate oscillator, the value of \hat{E}_g/\hat{E}_p becomes zero and no positive feedback is possible when the two voltages are exactly opposite in phase; oscillation ceases. Over a certain range the ratio \hat{E}_g/\hat{E}_p increases as the circuits are tuned to increase the phase difference between E_p and $-E_g$, and the path of operation opens out into an ellipse. In the Colpitts oscillator, Fig. 20.2, the excitation ratio is definitely determined by capacitors C_p and C_g and is very nearly given by

$$\frac{\hat{E}_g}{\hat{E}_p} = \frac{C_p}{C_g} \quad (20.4)$$

The two voltages E_p and E_g are very nearly opposite in phase, and the path is practically a straight line.

In the Colpitts circuit of Fig. 20.2,

$$P_{Lo} = \frac{\hat{I}_1^2}{2} \left(R_1 + \frac{\omega^2 M^2}{R_2} \right) \quad (20.5)$$

if the secondary circuit is tuned to resonance. Therefore, by (20.1), (20.3), and (20.5),

$$R_{Lo} = \frac{\hat{E}_p^2}{\hat{I}_1^2 \left(R_1 + \frac{\omega^2 M^2}{R_2} \right)} \quad (20.6)$$

$$\doteq \frac{1}{\omega^2 C_p^2 \left(R_1 + \frac{\omega^2 M^2}{R_2} \right)} \quad (20.7)$$

or

$$R_{Lo} \doteq \frac{L_1 \frac{C_g}{C_p + C_g}}{C_p \left(R_1 + \frac{\omega^2 M^2}{R_2} \right)} \quad (20.8)$$

The frequency of oscillation is very nearly that which makes the series reactance of the load circuit vanish. For the tank circuit alone, Fig. 20.2, the angular frequency of oscillation is approximately

$$\omega^2 \doteq \frac{1}{L_1 \frac{C_p C_g}{C_p + C_g}} = \frac{1}{L_1 C_1} \quad (20.9)$$

where C_1 is the capacitance equivalent to C_p and C_g in series. The frequency of oscillation may also be identified with the natural frequency of free oscillation of the circuit with negligible resistance.

When the secondary circuit is coupled to the tank circuit, the frequency of oscillation is nearly that which makes the series reactance of the coupled circuits, as viewed from the primary circuit, equal to zero, or approximately that of free oscillation of the coupled circuits. The wavelengths of free oscillation of coupled circuits¹ for one value of k are reproduced in curves *a* of Fig. 20.3. If the circuits were resistanceless, these curves would give the ratio λ_{osc}/λ_1 as the secondary circuit is tuned, *i.e.*, as λ_2 is varied. When λ_2/λ_1 is unity, the secondary circuit is in tune with the primary circuit but there would be two possible wavelengths of oscillation. When the resistance R_2 of the secondary circuit is taken into account, the condition of zero reactance of the system as viewed from the

¹ Fig. 4.1, Chap. VII.

primary circuit¹ gives the wavelength of oscillation. Curves *b*, *c*, *d*, Fig. 20.3, drawn for three values of Q_2 are the loci over which the reactance of the system as viewed from the oscillator tank circuit is zero and hence give the wavelength of oscillation. Curve *d* is for a low value of Q_2 and shows only a small effect upon the wavelength of oscillation. On the other hand, curve *b* is for a high value of Q_2 and shows a range of λ_2/λ_1 , in which there should be three possible wavelengths of oscillation. The dotted portion of this curve is unstable, however, so there are within this range of λ_2/λ_1 two

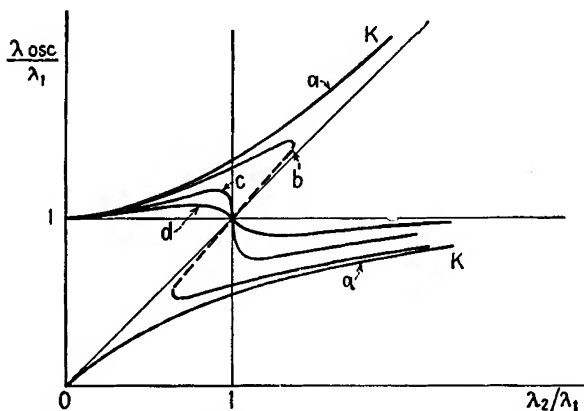


FIG. 20.3.—Wavelength of oscillation of coupled circuits.

possible wavelengths of oscillation. Curve *c* is a sort of critical case that divides those cases having one possible wavelength of oscillation from those having two. This critical case is identified by the relation

$$k = \frac{1}{Q_2} \quad (20.10)$$

When $k > 1/Q_2$, as for the locus shown in Fig. 20.4, and the secondary is tuned by increasing C_2 , that is, λ_2/λ_1 increasing, the wavelength of oscillation increases along the branch *jlm*. At *m* a sudden jump occurs to *n*. For a continued increase of λ_2/λ_1 , the branch *np* is traversed. If now λ_2/λ_1 is decreased, the course is *pnolj*. This hysteresislike loop gives rise to two sudden jumps in oscillation wavelength, often noticed when a tuned circuit is coupled to an oscillator. The loop is sometimes called the “drag loop.” The greatest amount of dragging and hence the greatest load on the

¹ Given by setting equal to zero the second term on the right-hand side of (7.2), Chap. VII.

oscillator occur as points m and o are approached. The positive feedback of the oscillator may be insufficient to maintain the system in oscillation up to points m or o , in which case the sudden jumps in oscillation wavelength occur before points m and o are reached, thus narrowing the range of λ_2/λ_1 over which there are two possible wavelengths of oscillation.

When an oscillator is used to feed power to an antenna or load, the drag loop should be avoided by fulfilling the relation

$$k < \frac{1}{Q_2} \quad (20.11)$$

by decreasing either k or Q_2 . Since the efficiency of power transfer from the oscillator tank circuit to the secondary circuit is given by

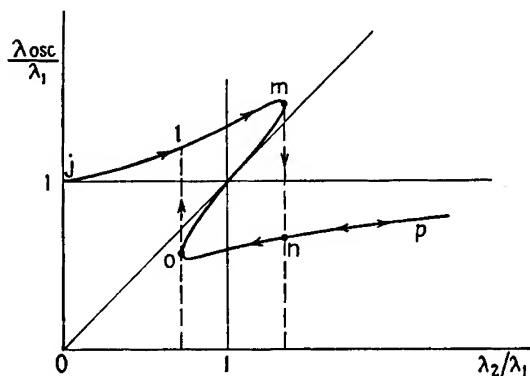


FIG. 20.4.—Wavelength of oscillation with $k > (1/Q_2)$.

(15.5), the additional condition of (20.11) may be substituted, giving as the efficiency of the circuits for an oscillator set, for a single wavelength of oscillation,

$$\eta_{cir} \leq \frac{Q_1/Q_2}{1 + \frac{Q_1}{Q_2}} \quad (\text{for oscillator}) \quad (20.12)$$

In this relation Q_1 is $\omega L_1/R_1$ and Q_2 is $\omega L_2/R_2$ at the frequency of oscillation; hence (20.12) can be written

$$\eta_{cir} \leq \frac{L_1 R_2}{L_1 R_2 + L_2 R_1} \quad (20.13)$$

From (20.12) it is apparent that, for good circuit efficiency, Q_2 must be small in comparison with Q_1 . It is difficult to feed power efficiently from an oscillator into a high- Q secondary circuit.

When (20.11) is fulfilled, the greatest transfer of power occurs when $\lambda_2/\lambda_1 = 1$ and the secondary circuit introduces only a resistance into the oscillator circuit. Furthermore, since $\lambda_{osc}/\lambda_1 = 1$, the secondary circuit does not alter the wavelength of oscillation.

21. Secondary Emission from the Grid.—It is explained in Sec. 9 that secondary emission from the grid of a tube decreases the grid current and may result in a reversal of this current. Secondary emission in a power amplifier decreases the driving power. Figure 10.3 shows that, the lower the value of R_L , the higher is the path of operation and the deeper it penetrates into the secondary-emission region and hence the greater are the effects of secondary emission.

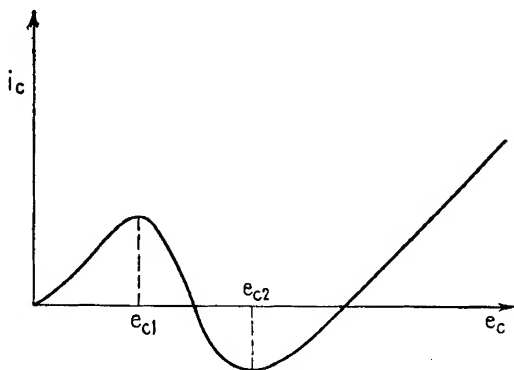


FIG. 21.1.—Possible variation in instantaneous grid current i_c during a radio-frequency cycle.

If the grid current during a cycle of the radio-frequency grid voltage is plotted, a curve of the type shown in Fig. 21.1 may result. For grid voltages between e_{c1} and e_{c2} a negative variational resistance is offered to the grid circuit. This negative resistance often sets up parasitic oscillations having a frequency high enough to permit a number of oscillations to occur during the time the grid voltage falls within the range $e_{c1} - e_{c2}$. These parasitic oscillations usually involve the inductance and capacitance of the leads from the grid coil to the grid or between tubes that are operated in parallel. The parasitic oscillations usually may be suppressed by including damping resistances in the grid connections to the tube.

Secondary emission in any tube in which the grid bias is produced by the voltage drop of \bar{I}_c through a resistance may seriously affect this bias voltage. An increasing secondary emission causes a decreasing bias voltage. The tube, if operating in parallel with others, takes a disproportionate part of the load, and its plate

becomes overheated. This decrease in negative bias voltage may progress so far as to "block" the tube. This blocking occurs when the grid current reverses, thus providing a self-maintained positive grid bias. The resulting excessive plate current is often so great as to destroy the tube before the plate circuit is interrupted by circuit breakers or fuses. Blocking is more apt to occur in oscillators and modulators than in ordinary Class *C* amplifiers, and the tendency to block is accentuated by decreasing the plate-load resistance R_L .

If secondary emission is caused by the condensation of active elements on the grid, the grid can be "decontaminated" by heating it to a temperature sufficiently high to evaporate the active elements. This can be done best by connecting the grid to the plate voltage supply and cautiously raising the voltage to obtain the necessary grid dissipation. A few trials may be necessary, the tube being put back in service after each trial in order to determine whether or not the contamination persists as shown by the grid current \bar{I}_c . The grids of some tubes are plated with graphite or other elements or are made of special materials that give small secondary emission.

22. Wide-band Power Converters, Nonsinusoidal Operation.—

In Sec. 5, power converters are divided into two general classifications according to the frequency band over which they operate. The wide-band converter operates into a load that offers a high impedance to a wide range of frequencies and is often a nonreactive load. For example, a modulator tube such as T_M in Fig. 19.1*b* supplies power at modulation frequency to the plate circuit of the power amplifier. The load characteristic of the plate circuit of the amplifier is nonreactive, at least for modulation frequencies that are low in comparison with the carrier frequency. The resistance of the amplifier as a load on the modulator may not be independent of the current. Such a load may have a characteristic of the type shown in curve *a*, Fig. 22.1. If the amplifier acted as a constant resistance, its characteristic curve would be a straight line illustrated by *b*. An oscillator may be modulated and hence be the plate load of a modulator tube.

The characteristic of the oscillator as a load in the modulator plate circuit might have somewhat the shape of curve *c*, Fig. 22.1. The characteristics are plotted in Fig. 22.1 in terms of \bar{E}_b and \bar{I}_b , the values of voltage and current averaged over a radio-frequency cycle. The variations at radio frequency are disregarded in considering the action of load in the plate circuit of the modulator tube,

where the variations occur at modulation frequency, since the r-f current is kept out of the modulator tube circuit by means of r-f choke coils or by-pass capacitors.

If the modulated amplifier or oscillator is placed in series connection or its equivalent (for example, by using a transformer M ,

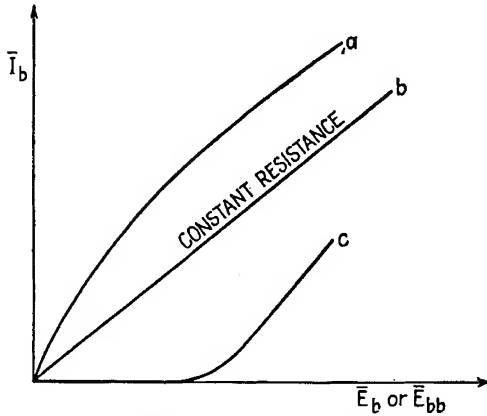


FIG. 22.1.—Variations in average plate current resulting from variations in plate voltage of a Class *C* amplifier or oscillator.

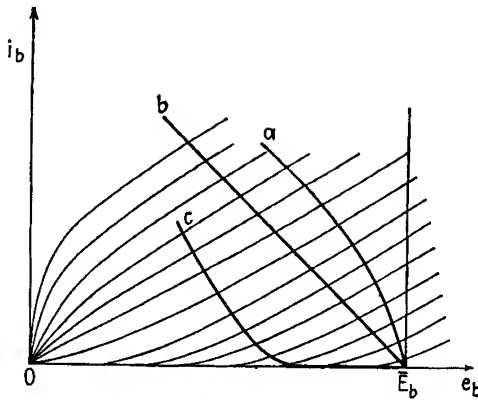


FIG. 22.2.—Load characteristics of Fig. 22.1 reversed and superimposed as load lines on the i_b - e_b diagram of a modulator tube.

Fig. 19.1a), in the plate circuit of the modulator, the mode of operation of the modulator can be determined by drawing the load characteristics of Fig. 22.1 reflected in the \bar{I}_b axis on the i_b - e_b diagram of the modulator tube, Fig. 22.2.

23. Frequency Multipliers.—Frequency multipliers are used in frequency-modulation and amplitude-modulation systems when the

radiated wave is of a higher frequency than can easily be stabilized, for example, by a quartz crystal. A frequency multiplier is simply a power-amplifier stage in which the plate-load circuit is tuned to a harmonic of the alternating grid voltage.

Frequency-multiplying stages usually do not require neutralization. The voltage across the plate tank circuit at the fundamental or grid input frequency is usually very small, and consequently the power fed back into the input circuit is insufficient to sustain oscillations at the input frequency.

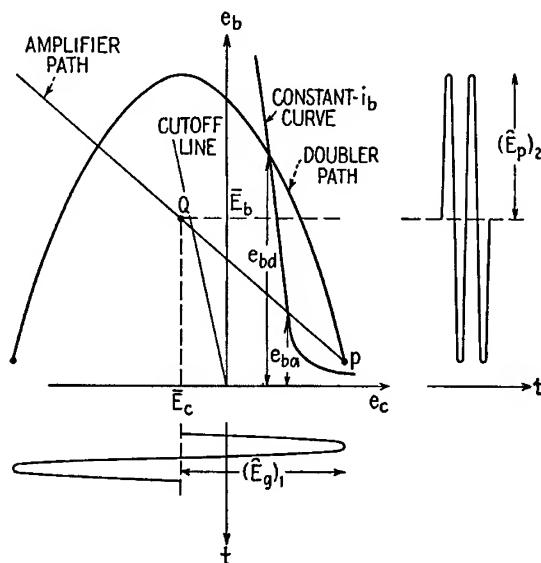


FIG. 23.1.—Paths of operation of a Class C amplifier and of a Class C doubler superimposed on the e_b - e_c diagram.

The operation of a frequency multiplier can be studied by using the e_b - e_c static curves of the tube, but instead of the path of operation being a straight line it is a parabola for a frequency doubler, an S curve for a frequency tripler, and other Lissajous figures for higher frequency multipliers.

The path for a doubler is shown in Fig. 23.1. The path of operation does not pass through the Q point as for a simple amplifier but is a single-valued path in the form of a parabola when the plate tank circuit has unity power factor for the second harmonic. Under this condition the average plate current \bar{I}_b is a *minimum*. When the tank circuit is detuned, the path opens into a double-valued path and the plate current *increases*. The highest tube efficiency, as in a

simple amplifier, occurs when the tank circuit has unity power factor for the output frequency.

The plate dissipation for the doubler is much greater than for a simple amplifier having the same quiescent and end point in its path of operation because the plate voltage e_b is much higher for the same plate current, as is seen by comparing e_{ba} and e_{bd} , Fig. 23.1.

The efficiency of the doubler can be improved by using a high negative grid bias and a correspondingly high driving voltage. The price of high plate efficiency is, however, a higher driving power. A frequency doubler inherently has a lower plate-circuit efficiency than a straight amplifier and therefore is confined usually to the low-power stages. Because the operating path of a doubler is in general higher on the i_b - e_b diagram than the path of an amplifier, the path of a doubler passes more deeply into the region of secondary emission from the grid, causing an accentuation of parasitics and other phenomena resulting from secondary emission.

CHAPTER XV

OSCILLATORS

1. Introduction.—An oscillator is a device for obtaining a periodically varying current from a steady power source. This current may have many different waveforms, depending upon the device used to produce the oscillations. A purely sinusoidal waveform may be desired, or special waveforms may be employed for special purposes, such as a saw-tooth wave for the sweep circuit of a cathode-ray oscillograph. This chapter is concerned with oscillators for the production of a fairly pure sine waveform, employing vacuum tubes and associated circuits.

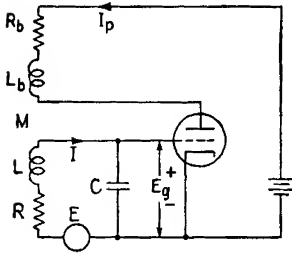


FIG. 2.1.—Regenerative tuned-grid circuit.

If the tube were removed, the alternating current I would be given by an expression such as

$$I = \frac{E}{R + jX} \quad (2.1)$$

where R is the resistance of the tuned circuit in the grid and X is the reactance of L and C in series, given by

$$X = \omega L - \frac{1}{\omega C} \quad (2.2)$$

With the tube replaced, there would exist an alternating voltage E from grid to cathode, given by

$$E_g = \frac{I}{j\omega C} \quad (2.3)$$

This would cause a variational plate current I_p to flow in the plate circuit. The current I_p flowing through L_b would, by means of the mutual inductance M , induce a voltage in coil L . This voltage would either aid or oppose the voltage E . Regenerative coupling exists when the connections are such that the induced voltage from L_b aids E . Let the coupling be regenerative, and let $|M|$ denote the magnitude of the mutual inductance¹ between L_b and L .

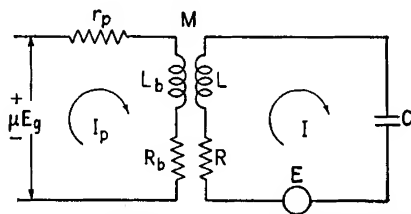


FIG. 2.2.—Circuit equivalent to that of Fig. 2.1 for small voltages.

Equivalent Circuit.—The equivalent circuit when E is small is shown in Fig. 2.2. Kirchhoff's laws applied to this circuit yield the following equations:

$$I_p(r_p + R_b + j\omega L_b) - j\omega|M|I = \mu E_g \quad (2.4)$$

for the left-hand mesh, and

$$-j\omega|M|I_p + \left(R + j\omega L - j\frac{1}{\omega C}\right)I = E \quad (2.5)$$

for the right-hand mesh.

The solutions to these equations may be placed in the form

$$I = \frac{E}{R_{eq} + jX_{eq}} \quad (2.6)$$

where R_{eq} and X_{eq} are the equivalent resistance and reactance of the circuit connected to the generator E , the regeneration produced by L_b being taken into account. R_{eq} and X_{eq} are given by

$$R_{eq} = R - \left[\frac{\mu \frac{|M|}{C} - \omega^2 M^2}{(r_p + R_b)^2 + \omega^2 L_b^2} \right] (r_p + R_b) \quad (2.7)$$

¹ By the conventions of Chap. V on Networks and Impedance Matching, M is here a negative number since, for the assumed positive directions of I and I_p , the magnetic fields would oppose each other. However, for clarity in the following discussion the symbol $|M|$ as used here represents only the magnitude of the mutual inductance.

and

$$X_{eq} = \omega L_{eq} - \frac{1}{\omega C} \quad (2.8)$$

where L_{eq} is an equivalent L given by

$$L_{eq} = L + \left[\frac{\mu \frac{|M|}{C} - \omega^2 M^2}{(r_p + R_b)^2 + \omega^2 L_b^2} \right] L_b \quad (2.9)$$

When $\mu|M|/C$ is greater than $\omega^2 M^2$, the effect of the regeneration is to reduce the resistance and to increase the inductance. Although the quantity in the brackets of (2.7) is small, the factor $(r_p + R_b)$ is large. Therefore the reduction in the resistance may be large. In (2.9) the quantity in the brackets is the same as that in (2.7) and is again small. The inductance L_b is ordinarily small, so that the increase in inductance is small. This increase in inductance produces a slight detuning effect. The principal effect of the regeneration is the reduction in the resistance. This assumes that any mechanical displacements introduced for the purpose of changing M do not result in any appreciable alteration of stray capacitance across the tuning coil.

Oscillation.—By increasing $|M|$, R_{eq} of (2.7) may be made equal to zero. Also, E may be of such frequency that X_{eq} of (2.8) is also zero. Then the ratio of I to E as given by (2.6) approaches infinity. Practically, this means that E may be withdrawn and the current I will be maintained by the action of the circuit. The circuit will “oscillate,” *i.e.*, produce a periodically varying current with only steady power sources in the circuit.

The natural frequency of free oscillation is the frequency at which R_{eq} and X_{eq} are zero. This frequency is given by

$$\omega^2 = \frac{1}{LC \left(1 + \frac{Q_p}{Q} \right)}$$

where $Q = \omega L/R$ and $Q_p = \omega L_b/(r_p + R_b)$. It is slightly less than the natural frequency of the LC circuit alone, because the equivalent inductance L_{eq} in (2.9) is slightly greater than the actual inductance L which appears in (2.2). This difference is not of great practical importance, for it can easily be counteracted by adjustment of C .

The ratio of I/E is plotted against the value $\omega|M|$ in Fig. 2.3. There are two values for which I/E approaches infinity. Between

these values lies the region in which the circuit will oscillate. If $\omega|M|$ is too small or too large, the circuit will not oscillate. In Fig. 2.3 the region to the right of the upper critical value of $\omega|M|$ is the region in which $\omega^2 M^2$ is greater than $\mu|M|/C$ in (2.7) and (2.9). This means that the coupling between L_b and L is so tight that the L_b circuit is coupling a large resistance into the L circuit. Oscillation cannot then occur since R_{eq} must be zero for oscillations to occur.

The larger critical value of $\omega|M|$ is seldom obtained in practice, since $|M|$ cannot exceed $\sqrt{L_b L}$ in magnitude and, in the usual design, L_b is much smaller than L .

The voltage E has been introduced in this discussion to show that under appropriate conditions the oscillatory current is self-sustaining. Practical oscillators must be self-starting as well as self-

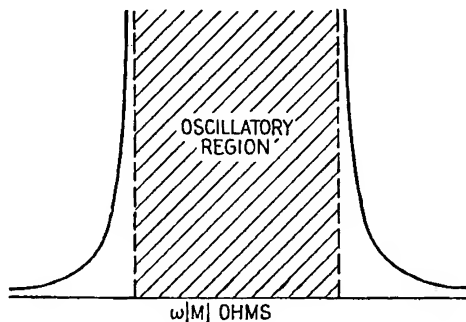


FIG. 2.3.—Variation of current in regenerative tuned-grid circuit.

sustaining. One might ask how the oscillator can be self-starting. The answer is that random fluctuations in voltage are always present in a circuit or tube, caused by thermal agitation in the circuit and tube, fluctuations in tube current due to shot effect, contact differences of potential, etc. These random voltages cover the whole frequency spectrum; but since the circuit is tuned, the circuit selects the component to which it naturally responds and amplifies this component strongly, owing to regeneration, so that the oscillatory condition is established. The oscillator is self-starting if, in its quiescent condition, the circuit and tube can select and amplify a suitable component of the random voltages existing in the circuit and if the regeneration is such that the output of the tube feeds back to the input a sufficient power to sustain the amplified current.

Intensity of Oscillation.—In the ordinary case (where the upper critical value of $\omega|M|$ cannot be obtained), the intensity of oscillation increases with increasing $|M|$, sometimes passing through a maxi-

mum. It might be thought that, since I/E is infinite, the current I would be infinite. But the preceding circuit theory was based on small amplitudes, for which the μ and r_p of the tube may be assumed constant. When the variational plate current is large, the instantaneous values of μ and r_p vary throughout the cycle. In general, this results in the inclusion of smaller values of μ and larger values of r_p than are experienced in small-signal operation with recommended bias voltages. The resultant decrease in the *average* value of μ , together with an increase in the *average* value of r_p , produces a progressive decrease in amplification, until finally the gain of the "amplifier" is barely sufficient to sustain the prevailing large-signal level, there being no remaining margin for further increasing the oscillation amplitude. The "grid-leak-bias" method of obtaining grid-bias voltage, Sec. 5, has a very stabilizing effect on the intensity of oscillation. In the final steady state the energy input from the power source in the plate circuit equals the energy dissipated in the tube and circuit.

3. Oscillator Circuits.—Except at the extremely high frequencies, practically all radio-frequency oscillators are essentially tuned amplifier circuits with means provided whereby some of the output is fed back into the input, the amplifier supplying its own signal.

The plate current increases in an ordinary amplifier when the signal drives the grid more positive with respect to the cathode. The increase in plate current causes a voltage drop in the load impedance, decreasing the voltage on the plate. This means that when the voltage on the grid goes up, the voltage on the plate goes down. There is a 180° phase difference between the variational voltage from grid to cathode and the variational voltage from plate to cathode. In an oscillating circuit in which the tube acts as an amplifier, this 180° relationship must exist. It is not necessary that the phase be exactly 180° , but it must be approximately 180° .

In an oscillating circuit there must be some form of energy storage, usually in the form of a tuned circuit. This is especially true when the oscillating amplitude is large. If the tube oscillates strongly, the current in the plate circuit may be cut off for part of the a-c cycle owing to the highly negative voltage applied to the grid during part of the cycle. Then the plate current flows in pulses, and in consequence the plate battery is supplying energy in pulses. Since the circuit is oscillating continuously, the circuit must provide some means of accepting the energy from the battery in pulses, and sending it out as a continuous alternating-current

wave. This function is performed by the tuned circuits, which can store energy in both the capacitor and the coil. The tuned LC circuit is often called the "tank circuit" since it is the place in which the energy is stored.

The tank circuit and the tube with its plate battery may with fair accuracy be compared to a flywheel and a one-cylinder engine firing once each revolution. The engine releases a pulse of mechanical energy each revolution. The tube and battery release a pulse of electrical energy each cycle. The flywheel stores and releases the mechanical energy, making the angular motion fairly uniform. The tank circuit stores and releases the electrical energy, making the voltage output a fairly good sine wave.

In the circuits that follow, the energy storage is supplied by an LC combination of some form. The tube is connected to the circuit in such a way that, when the alternating grid-to-cathode voltage is positive, the alternating plate-to-cathode voltage is negative. These features are essential.

The steady polarizing voltages may be applied to the grid and plate in several ways, to be discussed later.

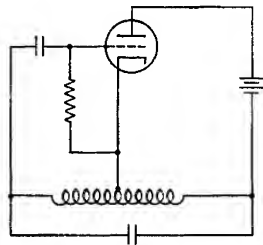


FIG. 3.1.—Hartley oscillator circuit.

Hartley Circuit.—The Hartley circuit, Fig. 3.1, employs a simple LC tuned circuit. The 180° phase relationship is obtained by connecting the grid to one end of the coil, the plate to the other end, and the cathode to a point somewhere in the middle of the coil. In some practical forms of this circuit, means are provided for connecting plate, grid, and cathode to different turns of the coil by means of clamps.

This circuit is a very good oscillator, since adjustments are not critical. The connections to the coil are such that the plate-to-cathode voltage is about μ times the grid-to-cathode voltage. There is one disadvantage. In any strongly oscillating circuit, the plate and grid currents are nonsinusoidal and therefore contain harmonics. The Hartley circuit provides no low-impedance path back to the cathode for these harmonic currents. For this reason the harmonic voltages across the tank circuit are somewhat larger in the Hartley oscillator than in some other types. The circuit as a whole is a "sure-fire" oscillator, in which the oscillating condition is readily established even when the circuit constants are not correctly chosen. With the circuit already oscillating it is often possible to improve the

amplitude and frequency by trial-and-error methods, when facilities for a more scientific design are lacking.

Colpitts Circuit.—The Colpitts oscillator, Fig. 3.2, connects the

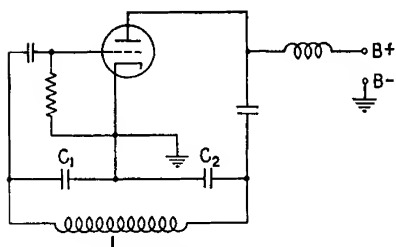


FIG. 3.2.—Colpitts oscillator circuit.

tube across the capacitive side of the LC circuit rather than across the inductive side as described above. The voltages applied to plate and grid are determined by the settings of C_1 and C_2 . The capacitance of C_1 and C_2 in series resonates with L to determine the frequency.

This circuit is somewhat more critical than the Hartley. It is not so easy to adjust, since every time C_1 is changed, C_2 must *also* be changed to keep the frequency constant. It has the advantage that the waveform is somewhat better than that for the Hartley since C_1 and C_2 provide a low-impedance path for the harmonic currents.

Figure 3.3 shows another form of the Colpitts circuit. The capacitance of the antenna takes the place of C_2 , Fig. 3.2.

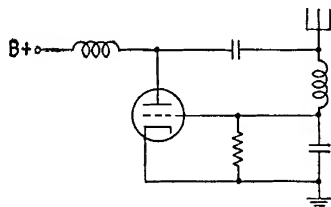


FIG. 3.3.—Oscillator circuit based on Colpitts circuit.

Meissner Circuit.—In the Meissner circuit, Fig. 3.4, there is no conductive connection between the tank circuit and the tube. The tank-circuit inductance is coupled to the grid circuit through inductance L_c and to the plate circuit through inductance L_b . In the Meissner circuit, L_c and L_b must not be coupled to each other. The coupling of L_c and L_b to the tank circuit is such that the required 180° phase relationship is attained.

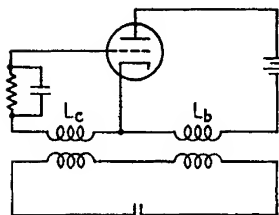


FIG. 3.4.—Meissner oscillator circuit.

Tuned-grid Circuit.—The “tuned-grid circuit,” Fig. 3.5 (the same as that of Fig. 2.1) is so called because the tuned circuit is located in the grid circuit.

Tuned-plate Circuit.—In this arrangement, Fig. 3.6, the tuned circuit is located in series with the plate and inductive coupling provided to excite the grid.

Tuned-grid Tuned-plate Circuit.—In this circuit, Fig. 3.7, there is a tuned circuit in series with the grid and a tuned circuit in series with the plate. There should be no magnetic coupling between the

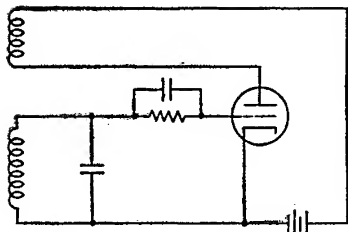


FIG. 3.5.—Tuned-grid oscillator circuit.

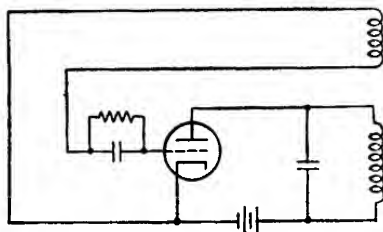


FIG. 3.6.—Tuned-plate oscillator circuit.

grid and plate coils. The plate circuit is coupled to the grid circuit by means of the capacitance between the plate and the grid of the tube. If the plate load is inductive at the frequency of oscillation,

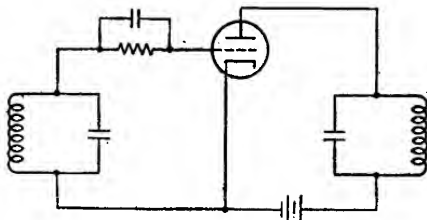


FIG. 3.7.—Tuned-grid tuned-plate oscillator circuit.

then, with proper circuit constants, enough energy is fed back to the grid circuit through the grid-to-plate capacitance to maintain oscillations. The tuned-grid circuit is sometimes replaced by a self-resonant choke coil.

4. Practical Considerations, Plate-voltage Supply.—The steady plate voltage may be fed to the tube in two ways, known as series feed and shunt feed.

Series Feed.—In the series-feed method, Fig. 4.1, the plate-supply battery is connected in series between the plate and cathode of the tube. A typical case of this is shown in Fig. 3.1 for the Hartley circuit. The disadvantage of the plate-battery connection of Fig. 3.1 is that the battery is at a point in the

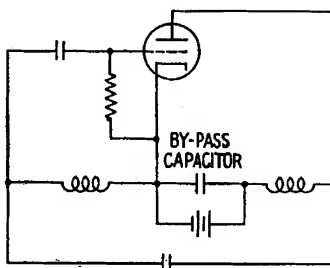


FIG. 4.1.—Series feed with battery at cathode potential.

circuit which has high radio-frequency potential with respect to the cathode. The plate power source is usually physically large and has a large distributed capacitance to ground. In Fig. 3.1 this capacitance is across part of the tank coil and in some cases will seriously interfere with the tuning of the circuit.

This objection is removed by the connection shown in Fig. 4.1, where the battery has been placed at cathode potential by splitting the coil. A capacitor is shown connected across the battery to by-pass the large radio-frequency currents existing in the tank circuit. This connection introduces the disadvantage of placing part of the coil at high d-c potential with respect to ground, and the tuning capacitor now has to withstand the sum of both d-c and r-f voltages.

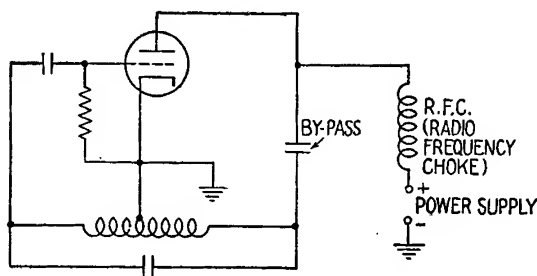


FIG. 4.2.—Hartley oscillator circuit with shunt feed.

The series feed is used chiefly in low-power circuits where the voltages are low. In practice, the d-c power supply is by-passed for r-f current by a capacitor.

Shunt Feed.—In the shunt-feed method, Fig. 4.2, the plate-supply battery is connected in shunt between plate and cathode, as shown in Fig. 3.2. To prevent the battery from short-circuiting the plate to cathode at r-f potential, a choke coil is connected in series with the battery. This choke coil is an inductance that has a low d-c resistance and a high radio-frequency impedance. Sometimes the self-capacitance of the choke coil is used to resonate the coil so that the r-f impedance is greater than the ωL value for the coil.

The shunt-feed circuit must be used where there is no d-c path from plate to cathode in the oscillatory circuit, as is the case for the Colpitts circuit. At other times, the shunt-feed circuit is used to keep the plate supply at ground potential. Also, a capacitor must be used to prevent the tank coil from short-circuiting the d-c plate supply to the cathode. This capacitor is made sufficiently large to

offer negligible impedance to the r-f plate current and must have sufficient insulation to withstand the sum of the d-c and the r-f plate potentials.

The Hartley circuit of Fig. 3.1 is redrawn in Fig. 4.2, showing the use of shunt feed.

5. Practical Considerations, Grid-voltage Supply.—Except with very low-power oscillators, a battery cannot be used as a grid-bias supply. High-power oscillators require a high enough grid bias to operate the tube Class C. If this value of bias were supplied by a battery, no plate current would flow when the plate-supply switch was closed, and oscillations could not start. At the beginning the bias must be small to allow plate current to flow. Under running conditions, the bias must be large to prevent too much plate current or excessive grid current. This problem is solved by the use of what is called grid-leak bias or grid-resistor bias.

Grid-leak Bias.—In Fig. 4.2 a resistor is shown connected from grid to cathode. The d-c grid current must return to the cathode through this resistor. In so doing the d-c IR drop through this resistance places the grid at a negative d-c voltage with respect to the cathode.

Before oscillations start, there is no voltage applied to the grid, and hence no grid current. The bias is zero, so that plate current exists and oscillations start. When oscillations start, r-f voltage is applied to the grid, causing the grid to draw current. The average value of this current times the value of the grid resistor gives the value of the grid-bias voltage.

The grid resistor has a very beneficial effect in controlling the oscillation. If the intensity of oscillation becomes greater, the r-f voltage applied to the grid is greater and the grid current is greater. The grid bias is increased, causing the plate current to decrease, which in turn cuts down the intensity of oscillation. If the intensity of oscillation becomes smaller, the grid current becomes smaller. The bias is decreased, allowing more plate current to flow so that the oscillation level may return to normal. Therefore, the use of grid-leak bias imparts a stability to the intensity of oscillation since any increase or decrease in amplitude of oscillation causes a counter-acting increase or decrease in grid bias.

In some cases the grid resistor is in series with the grid. In such cases the resistor is by-passed for radio frequencies by a capacitor. A typical case of this is shown in Fig. 3.5.

In all cases where grid bias is obtained by means of a resistor, a

capacitor in some form is used with it. The capacitor by-passes the resistor for r-f currents, as in Fig. 3.5, or forces the d-c grid current to flow through the resistor, as in Fig. 4.2.

Intermittent Oscillation.—If the time constant of the grid-leak and grid-capacitor combination is too large, the oscillation may start and stop periodically. This is called intermittent oscillation. Such action may be analyzed as follows: When the plate-supply switch is closed, oscillations start and build up to a high intensity. A high voltage is applied to the grid, causing a large grid current. This current produces a high d-c voltage across the grid-leak resistor; to this voltage the grid capacitor is charged. This decreases the intensity of oscillation. If the time constant of the resistor-capacitor combination is so high that the capacitor cannot discharge rapidly, the grid bias will stay at a high value even though the oscillation is decreasing. The controlling action described above does not now operate rapidly enough, and oscillations die out completely. Even after oscillations cease, the plate current will remain at a very low or zero value because of the high negative bias that is retained on the grid. Oscillations will not start again until the capacitor has discharged sufficiently to allow the grid bias to become low enough so that the amplifying action of the tube may start oscillations again.

The cycle of starting and stopping repeats itself. With a large product CR , the period may be very long. The intermittent oscillator is sometimes used to time sweep circuits, the interruption frequency determining the duration of the sweep cycle.

Positive-grid Blocking.—If for some reason the grid current should be reversed, the d-c current through the grid resistor would cause the grid to be biased positively with respect to the cathode. This reversal of grid-current flow may be caused by secondary emission from the grid or by thermal emission if the grid becomes red-hot owing to excessive dissipation. Such conditions are caused by too high a voltage applied to the grid, due to a surge or to a high grid driving voltage. If the grid-leak resistance is too high, the grid may become so positive that the reversed grid current is steadily maintained and the grid acquires a *stable* positive bias. This condition is independent of the value of the grid capacitor. With the grid bias maintained positive, the plate current increases, in some cases to the saturation value. This will stop the amplifying action of the tube and may cause the tube elements to melt. Circuit breakers or fuses in the plate circuit are designed to prevent such damage but

are often too sluggish to prevent destruction. The cessation of tube activity caused by grid-current reversal is called "positive-grid blocking."

Sometimes an oscillator having intermittent oscillations is also called a "blocking" oscillator since, during the time when there are no oscillations, plate current is prevented, or "blocked," by the high negative grid bias. The more correct name is the term "intermittent oscillator" as used above.

6. Practical Considerations, Power Output.—In presenting the basic triode circuits, Sec. 3, no reference was made to specific measures for obtaining useful power output, for no one method is universally accepted. The choice depends upon the application. Again taking the Hartley circuit as a representative illustration, one

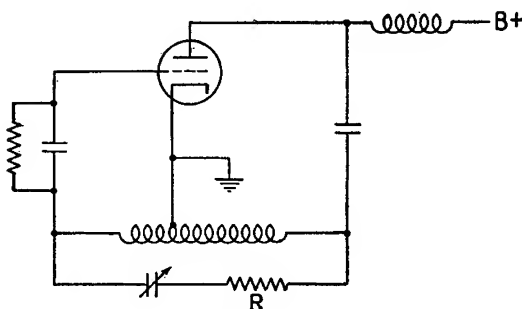


FIG. 6.1.—Output obtained by loading the tank circuit.

may begin by considering a load resistor R , placed in series in the tank circuit, Fig. 6.1. R is any device that is to be heated by alternating current generated by the oscillator. (For example, the filament of a cyclotron, exposed to a strong magnetic field, may be distorted by d-c heating current or vibrated excessively by 60-cycle alternating current. At 2,000 cps, however, the vibration should be inappreciable.) Frequently R is not a separate component, visible to the eye, but represents, instead, power usefully converted into heat in the field of a capacitor or an inductor. Many industrial applications are of this character. In the bonding of plywood, stacks of wood to be cemented are placed between the parallel plates of a large capacitor. In heating sheet steel in a modern tin-plate mill, high-frequency magnetic induction is employed. In high-frequency therapy an artificial fever is induced and controlled by placing the patient in the field of a coil or a capacitor.

One may generalize still further. R may be the radiation

resistance of an antenna or the input resistance of a transmission line, leading to an antenna. R may be the equivalent primary resistance of a loaded transformer, with secondary connected to an antenna, to a line, or to any circuit that dissipates power. R may be the series resistance, equivalent to a shunt load bridged across the entire inductor or any part thereof.

No matter where the power goes ultimately, R determines the I^2R rate at which power is to be transferred out of the tank circuit by current of given amplitude. Hence the very best "flywheel effect" is obtained by making R as small as possible, any load that may be applied necessarily reducing the Q of the tank circuit and reacting unfavorably upon the waveform. Since the frequency of oscillation depends slightly upon R , and not exclusively upon the LC product, any alteration of load will be accompanied by a slight shift of frequency. In industrial and therapeutic applications such deviations of waveforms and frequency are likely to be of little consequence. In radio communication, particularly in the crowded portions of the radio spectrum, excellence of waveform and stability of frequency are very often matters of major concern. Waveform distortion means the production of radio-frequency harmonics. If these reach the antenna, they are likely to cause illegal radiation on channels assigned to other services. Frequency deviation also can be a serious problem, errors as small as 0.002 per cent often being undesirable. Intentional modulation of amplitude, in the transmission of speech or music, constitutes an effective fluctuation of load accompanied by an undesired frequency modulation. The resulting wobble distorts the received signal. Since these disadvantages are fundamental, improvement of waveform and frequency stability can be secured only by operating the oscillator under light and constant load.

Buffer Amplifier.—The limitations mentioned in the preceding paragraph do not preclude obtaining a considerable *voltage* from the oscillator. However, the driven circuit should demand relatively little *current* from the source and should therefore require only a small amount of power. Furthermore, it should be a truly one-way circuit, permitting little or no reaction on the oscillator from any external cause. At low frequencies an ordinary triode amplifier satisfies these specifications. At radio frequencies the grid-to-plate capacitance causes feedback, and the prevailing phase relationships often result in a very considerable lowering of the effective grid-to-cathode resistance. This may be prevented by careful neutraliza-

tion, but it is usually simpler to use a pentode or tetrode tube as the amplifier, Fig. 6.2, thus minimizing internal feedback. A cathode follower may also be used. An amplifier of any type, used for this purpose, is called a "buffer amplifier" or simply a "buffer." It is intended primarily as a one-way coupling device, and any voltage gain secured in the buffer stage is merely incidental. Sometimes the buffer serves also as a frequency doubler or tripler. As such, it

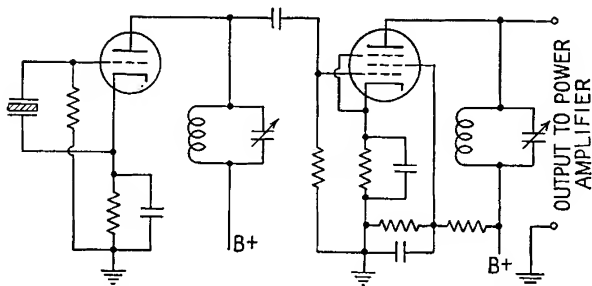


FIG. 6.2.—Use of buffer amplifier to isolate oscillator.

selects the desired harmonic from the harmonics already present in the oscillator circuit or else intentionally distorts the wave by nonlinear amplification.

Electron-coupled Oscillator.—By the use of multielectrode tubes, the functions of the oscillator and the buffer often are combined within a single envelope. For example, in Fig. 6.3 the three inner-

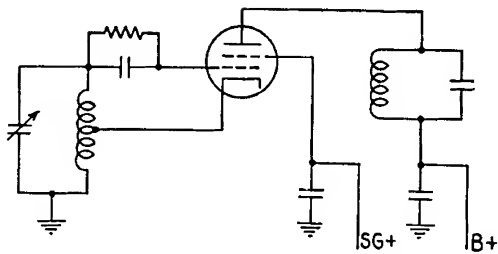


FIG. 6.3.—Electron-coupled oscillator.

most electrodes are connected in a Hartley circuit, the screen filling the role of the usual triode plate. Contrary to ordinary oscillator practice, the cathode of this tube is *not* grounded, but the screen is grounded by a capacitor of negligible impedance. Since this implies existence of an r-f voltage from cathode shell to ground, it is desirable to select a tube having a low capacitance from cathode shell to heater wires. The screen grid captures enough electrons

to maintain the oscillation, but many additional electrons pass through the spaces between the grid wires. These electrons continue on their course, most of them arriving ultimately at the plate. In fact, the outer portion of the tube may be considered as a diode, the grounded screen grid serving as an equivalent cathode (an apparent source of electrons). However, unlike the ordinary diode, the electrons from this virtual cathode are projected in periodic spurts. In the external plate load, a resonant circuit, tuned to the oscillator frequency or to an appropriate harmonic, selects the desired a-c component to be used in driving successive power amplifiers. As the output coupling is provided by the electron stream, the circuit is referred to as an "electron-coupled oscillator." Assuming adequate external shielding, the grounded screen grid extends this shield through the tube itself, and effectively prevents reaction. The three inner electrodes may be connected in any of the basic triode circuits and may be crystal-controlled if desired.

7. Coupled-circuit Oscillators, Drag Loop.—In an effort to reduce the radiation of harmonics, or for other reasons, the output

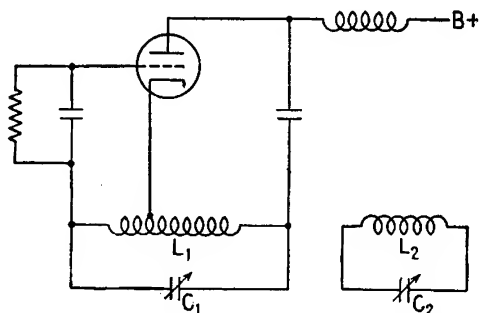


FIG. 7.1.—Tank circuit coupled to a second resonant circuit.

circuit of an oscillator sometimes includes two or more resonant circuits such as L_1C_1 and L_2C_2 , Fig. 7.1. With sufficiently close coupling, the interaction of the resonant circuits produces certain peculiarities that are likely to confuse an inexperienced operator. Partly because of these difficulties, such multiple-resonant circuits are seldom encountered nowadays in their obvious form with separate coils and capacitors. Very similar defects appear, however, in less patent form, whenever a single resonating structure, say a quartz crystal, or a tuning fork, or a hollow cavity, or a wave guide, exhibits a number of possible modes of vibration. Some-

times these modes interact as if separate coupled resonators were present. Accordingly, a brief explanation of coupled-circuit oscillators in general serves as an introduction to the discussion of crystal oscillators.

In the circuit of Fig. 7.1 let the dials of the capacitors C_1 and C_2 be calibrated in wavelength and be set initially at any common value, for example 300 m. By this we shall mean that circuit 1 has been calibrated in its position in the oscillator, but with circuit 2 far removed, or completely uncoupled, or greatly detuned. Similarly, circuit 2 has been calibrated as a separate isolated unit, removed from the influence of circuit 1. Now having brought these calibrated circuits into close proximity and set each pointer at 300 m, it is not to be supposed that the calibrations will still be valid and that the oscillator will emit a 1,000,000-cycle sinusoidal voltage wave. Since similar coupled circuits emit a wave of double frequency in transient oscillation and since the resonance curve of coupled circuits driven sinusoidally is double-humped, it might be expected at first that the oscillator would operate simultaneously on two frequencies, thus emitting a wave modulated at the beat frequency. Indeed, such action *can* occur in exceptional cases, though it does not *ordinarily* take place. One of the two wavelengths would be longer, the other shorter than the resonant wavelength, which in this example is 300 m.

More commonly, one of the two possible modes of oscillation builds up more rapidly than the other by reason of better phase angle and therefore greater gain. This quickly overpowers the less efficient mode, suppressing it entirely, and the oscillator operates at constant amplitude on *one* of the two possible oscillator frequencies. In the case of the coupled circuits, each nominally set at 300 m, the lower frequency mode ordinarily prevails and the oscillator starts in the long-wave condition, for example at 342 m, the actual departure from 300 m depending upon the coupling coefficient. Accordingly, operation begins at point 1, Fig. 7.2.

In an attempt to lower the frequency still further, hoping to attain a wavelength of 400 m, the operator increases capacitance at C_2 , as indicated by an increased value of the dial reading, λ_2 . In agreement, the operating point slides up the λ' branch of the wavelength curve, Fig. 7.2, say to 360 m at point 2. Having succeeded thus far, the operator advances the same dial farther. By this time the alteration of phase angle has proceeded so far that an entirely new start (opening and reclosing the plate power switch for

example) would favor the short-wave mode of oscillation represented by the λ'' branch of the curve. Nevertheless the λ' mode has a "vested interest" in the operation, as the λ' oscillation is already vigorously in progress and the grid is being driven by a large r-f voltage of frequency appropriate to λ' operation. The λ'' oscillation has no opportunity to start in the face of such pre-established competition. In this manner, by continuous sliding of the setting of C_2 the operator can "drag" a mode of oscillation, originally established under favorable circumstances, part way into a range of adjustment that normally favors the alternate mode. There is a limit, however, beyond which this dragging action cannot

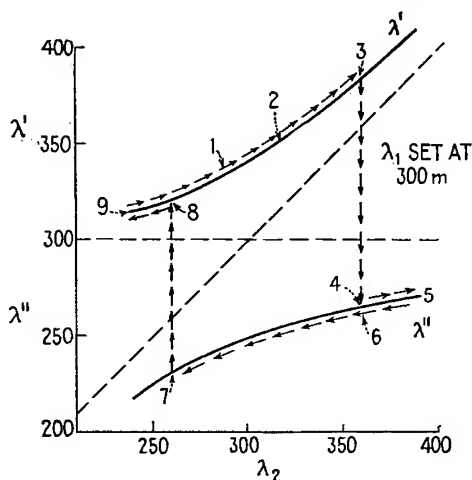


FIG. 7.2.—Example of drag loop.

be stretched. On reaching point 3 (say 384 m) the relative advantage of mode λ'' has increased to such an extent that it prevails in spite of the "vested interest" which mode λ' has acquired as a "going concern." As C_2 is gradually advanced, the frequency of oscillation suddenly jumps upward, the wavelength dropping vertically from the λ' branch to the λ'' branch. The attempt to obtain 400 m, by adjustment of C_2 alone, has failed. The wavelength has received a sudden setback, carrying it to 267 m, Fig. 7.2. Further continuous increase of C_2 will cause the wavelength to increase, but it now increases to point 5 from point 4 rather than from point 3 upward.

An inexperienced operator, puzzled by the apparently erratic jump of frequency, now tries to retrace his steps by gradually

decreasing C_2 , hoping to determine whether that frequency jump really did occur. To his further confusion it takes place this time at a different scale setting. In terms of the coupled-circuit diagram this is quite natural, however. Point 6 coincides with point 4, but the λ'' mode is now the "going concern" and is still favored. Hence no jump is to be expected as yet. On approaching point 7, however, the λ'' mode again becomes the favored mode, suppressed only because the λ'' oscillation is already established. At point 7 the oscillator can be "dragged" no further into the short-wave region by manipulation of C_2 alone. Hence at this point the operator observes a sudden drop of frequency, the wavelength shifting 90 m upward to the λ' curve. Any further decrease of scale setting, λ_2 , causes the wavelength to slide downward along the λ' curve from point 8 to point 9. If the capacitance again be increased, the oscillation can be brought back to its original condition 1. Continuous variation of C_2 , back and forth, causes the emitted wavelength to travel in a clockwise sense around the four-sided open figure 3, 4, 7, 8, which is termed a "drag loop."

8. Crystal Oscillator.—If a single crystal or a rectangular block properly cut from a single crystal be squeezed with a steady pressure, applied by insulated jaws of a vise, there is frequently observed a steady difference of potential between opposite faces of the block. Insulated tin-foil sheets or other metal plates adjacent to the crystal faces acquire a steady charge dependent upon the pressure. Such phenomena are called "piezoelectric" effects (pressure-electrical effects). The effect is reversible. When a crystal is placed in an electric field, slight changes of dimensions result. Thus a battery attached to the upper and lower plates of a parallel-plate capacitor slightly deforms a crystalline dielectric, no matter whether the capacitor plates are cemented to the crystal slab or are placed loosely above and below it. These results are not particularly surprising since all dielectric materials are considered to be in a strained condition owing to the relative displacement of positive and negative charge. In crystalline material the cumulative effect of the individual atomic distortions yields a barely measurable distortion of the block as a whole. Conversely, a sufficient mechanical distortion slightly displaces the positive and negative ions in the crystal framework, producing a systematic electrical strain throughout the block and in its immediate vicinity. The magnitude of the effect varies widely in different types of crystals. Rochelle-salt crystals are particularly vigorous but are excessively fragile.

Quartz offers fair sensitivity, together with considerable mechanical strength, which makes it desirable for most piezoelectric devices. Other crystalline materials, such as tourmaline, are occasionally used. Fused quartz is of no value, and even the quartz crystals must be very carefully selected. Very slight imperfections destroy electrical sensitivity.

The distortions produced by a steady electric field or a slowly varying field are almost imperceptible. However, by continuously increasing the frequency of the power source connected to the metal electrodes the applied alternating voltage eventually resonates with the mechanical vibration of the crystal. Thus a compressional wave, initiated by the electric field, travels back and forth between the parallel faces of the crystal wafer, reinforced at the completion of each cycle of travel by an accurately timed recurrence of the electric field that set it in motion. At this resonant frequency the amplitude of mechanical vibration increases notably, and the current drawn from the power source also shows a resonant change. In fact, the crystal may even be fractured by too violent mechanical vibration if the voltage applied to the plates is too high. The natural frequency of the compressional vibration is determined primarily by the velocity of sound in the quartz and by the thickness of the crystal slab. In the sizes normally obtainable the vibrating quartz resonates at a frequency far above audibility, thus emitting a supersonic "sound" wave that is propagated through the external medium, which is usually air, though depth-finding, antisubmarine measures and related applications sometimes involve crystals immersed in water. The stimulation of chemical reactions by agitation and the examination of imperfections within solid bodies by supersonic reflections provide representative industrial applications for such vibrators. The device also functions as a microphone, external sound vibrations causing crystal motion, with resultant induction of an electric field. If resonance is undesired, it may be checked by mechanical damping.

For investigating crystal applications in general it is convenient to represent the crystal, in its holder, by an equivalent electrical network. This may be reduced to the four-element network of Fig. 8.1. Hidden in a box with exposed binding posts A' and B' , this network theoretically could imitate the crystal by drawing the same current from the same voltage source at any given frequency. The capacitor C_1 provides for the action of the quartz as a simple dielectric enclosed between parallel capacitor plates. The series-

resonant circuit C_2LR provides for the additional current due to the *motion* of the quartz, important only at or very near the resonant frequency. This qualitative reasoning has been verified by a mathematical analysis of the effect of the crystal motion.

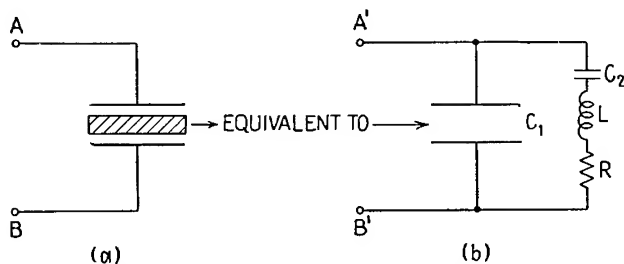


FIG. 8.1.—Equivalent circuit of crystal in its holder.

The conventional tuned-plate tuned-grid circuit, or TPTG oscillator, exhibited in Fig. 8.2a, may be modified slightly by splitting the grid-circuit tank capacitor into two series portions, thus arriving at Fig. 8.2b. In this diagram the elements RLC_2C_1 are arranged in the same relative configuration as in Fig. 8.1 and are

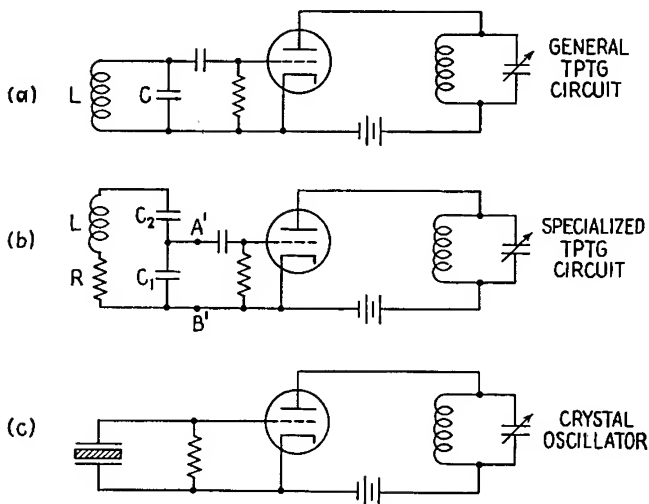


FIG. 8.2.—Development of a crystal-oscillator circuit.

equipped with the same pair of accessible binding posts $A'B'$. Hence the crystal may be substituted in place of its equivalent circuit. The grid capacitor is no longer necessary and may be discarded, though the grid leak is retained. A very common form

of the crystal oscillator thus becomes a special application of the familiar TPTG circuit. C_2C_1 serves as a capacitive potential divider. The voltage externally available across C_1 is very much smaller than the internal, inaccessible voltage theoretically existing across LR , but it has the same phase as the voltage across C and is, in magnitude, entirely adequate to drive the grid.

Since the crystal is equivalent to a tank circuit, it is apparent that it may be substituted in place of an LC circuit in several other oscillator circuits and also may be used as an element in filter design. In fact, it is an extraordinarily good tank circuit since the energy dissipation in mechanical friction in quartz is exceedingly low in comparison with the energy stored in the vibration. The circuit of Fig. 8.1*b* is "equivalent" to the crystal of Fig. 8.1*a* because we may draw such a circuit on paper, attach equivalent numerical values to the elements RLC_1C_2 , and may use the diagram in reasoning processes such as that presented in Fig. 8.2. However, we should not attempt to take our diagram into the shop and ask a mechanic to build the circuit according to these specifications. The relative numerical values of the circuit elements are absurd in terms of practical construction from existing copper, iron, permalloy, etc. The Q of the crystal (and its equivalent circuit) is likely to be at least ten times larger than the best electrical tank circuit that a shop can construct. Herein lies the main advantage of the crystal resonator and the crystal oscillator. As the Q increases, the frequency of oscillation is governed more and more by the properties of the tank circuit itself, and minor dependence upon amplification factor, plate resistance, etc., is progressively minimized. Hence, drifting of line voltage, aging of tubes and fluctuation of load have a reduced effect.

However, accidental or intentional alteration of the quartz block has an immediate effect upon frequency of oscillation. The crystal block is adjusted to the correct frequency by sawing it slightly oversize, then reducing thickness by grinding and polishing, using the best techniques that have been developed for the construction of precision lenses and similar optical devices. A tolerance of 1 cycle in 5,000,000 would not be considered exceptional. Where such precision is desired, the crystal is tested repeatedly in its permanent crystal holder, preferably in the oscillator in which it is to be used. After each test the flat surface is polished lightly with very fine rouge or similar abrasive material. (Larger tolerances should be specified wherever possible, as this permits mass-produc-

tion methods and greatly reduces the cost.) Great care must be taken to avoid removal of excess material, as this would destroy the usefulness of the crystal. A very minor overshoot can sometimes be corrected by grinding material from an edge, since a decrease of width or length has a second order effect upon the frequency of the desired compressional vibration.

Change of temperature alters the elasticity and density of the quartz as well as its external dimensions. In general, this results in a slight alteration in frequency. This may be considered as an advantage or a disadvantage, depending upon the method of operation. If the crystal is to be used in a carefully thermostated oven, control of the exact oven temperature permits a very minor adjustment of the precise frequency. By this method a precision of 1 part in 5,000,000 may be obtained and held. This is necessary in standard frequency devices. Simple and inexpensive ovens have often been used in transmitters. On the other hand, one often prefers to use a crystal at room temperature, accepting the changes of temperature that are then inevitable. In this case a very low temperature coefficient would be desired. Fortunately, the elastic properties of the quartz depend upon the exact angle of propagation within the crystal, relative to the axes of the original crystal matrix. These axes are seldom evident by inspection of the rough uncut quartz, but they may be determined with precision by a combination of optical and X-ray devices. Then, by exact control of the angle of the saw, one may obtain slices that have a positive or a negative temperature coefficient. For one particular angle, the coefficient is zero. This condition is approximated in many modern crystals.

Concealed flaws, or minute cracks, or the loss of a chip from an edge or corner may make a crystal respond poorly or not at all. A shift of position within the holder may cause a slight error in frequency. Crystals are easily destroyed by overheating or by excessive vibration. Fragility naturally increases as the thickness is decreased. Quartz crystals designed for frequencies of 6 or 7 meps are fairly common, and they can be ground for considerably higher frequencies. In general, however, this is unnecessary. Conservative practice usually dictates the use of doublers and triplers, employed in cascade if necessary, so that a sturdier low-frequency crystal may be installed.

A particularly annoying defect, common in many of the cheaper or older crystals, is the occurrence of drag-loop action. In addition to the desired compressional vibration, which has a frequency

determined primarily by the thickness of the quartz plate, compressional vibrations also may exist in the direction of the width or the length. Torsional and flextional vibrations are also possible. Normally, these undesired fundamental modes of vibration have natural frequencies far below the frequency of the "thickness vibration" and exert no appreciable effect. However, their harmonics are very numerous, and some harmonics are likely to be in the vicinity of the desired thickness vibration. On changing the operating temperature, for the purpose of making a minor adjustment in frequency, the effect shown in Fig. 8.3 may be obtained. In this plot, the zero of the frequency scale is far below the bottom of the page, the entire variation of frequency is relatively small, and yet within this narrow range the adjustment

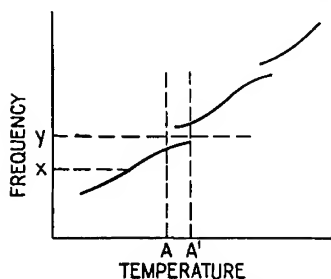


FIG. 8.3.—Drag-loop effect in a crystal vibrator.

appears to be erratic and jumpy. For temperatures within the band AA' the frequency curve is double-valued, one branch being followed as the temperature increases, the other branch as the temperature falls. A smooth and continuous variation of temperature eventually produces a sudden jump of frequency, the location of the jump depending upon the direction of temperature adjustment.

If the operator seeks to attain frequency x , he is indeed fortunate. Should the desired frequency fall at point y , the crystal is useless in its present form. Before consigning it to the wastebasket the experienced operator may grind a small amount of material from an edge or may even chip off a corner. By so doing he may happen to throw the interfering harmonics to a less objectionable position on the frequency-temperature scale. The entire curve will also be shifted in position by a slight amount, which may improve the situation or may simply aggravate the trouble.

The same defect can cause a similar discontinuity in the curve of frequency vs. thickness. Naturally the entire drag loop never appears, as there is no method for restoring quartz once removed. However, the existence of a discontinuous jump of frequency, perhaps encountered just as the desired frequency is approached, can cause considerable waste of time and effort and does cause numerous rejects.

Fortunately, this defect can be avoided. The mere existence

of a harmonic close to the frequency of the desired thickness vibration is not ordinarily preventable since the harmonics are very numerous. This would do no harm, however, if the two vibrations were not coupled. In general, deliberate excitation of the thickness vibration automatically produces the unwanted modes of vibration as well. For one particular angle of the crystal saw, relative to the axes of the rough crystal, this coupling does *not* take place. The critical angle that produces such decoupling does *not* coincide with the angle that produces zero temperature coefficient. However it happens that the difference in angle is not great, and a suitable engineering compromise is feasible. Hence one may now purchase a crystal with one advantage or the other or may secure a low temperature coefficient with approximate decoupling, making the irregularities very small.

It has recently been discovered that irradiation by X rays reduces the natural frequency of quartz crystals.¹ This process in some cases has eliminated much of the final fine grinding formerly necessary in adjusting the crystal to an exact frequency.

9. Other Electromechanical Oscillators, Magnetostriction.—The exceedingly good frequency stability of the crystal oscillator is obtained because the energy stored by elastic and inertial forces is large in comparison with the frictional dissipation of energy per cycle. However, this fundamental advantage is not a unique property of quartz crystals. Instead, such sharpness of resonance is exhibited by good mechanical vibrators in general. Stabilization of an audio-frequency vacuum-tube oscillator by means of a tuning fork is relatively easy and has been employed in certain standard-frequency systems and in precision speed control of various printer-telegraph devices. Plate current of the oscillator tube energizes an electromagnet and deflects the steel fork, the resultant motion generates an alternating voltage that drives the grid, and the oscillation is maintained. This may be regarded as a specialization of the Meissner circuit, the tuning fork taking the place of the equivalent electrical tank circuit.

Prof. G. W. Pierce, who invented the crystal oscillator, has developed a somewhat similar device that makes use of magnetostriction. Current from the plate circuit of an oscillator energizes a solenoid. The resultant magnetic field controls the magnetization

¹ X-rays Lower the Frequency of Quartz Plate, *Electronics*, **17** (11), 166, November, 1944.

of one end of a nickel rod, which projects through the solenoid, Fig. 9.1.

Alteration of magnetization produces a slight change in the physical dimensions of the metal by the cumulative effect of a rotation of innumerable atomic groups that act like tiny compass needles. As a result, a compressional wave is propagated down the rod, traveling as a sound wave within the metal. On reaching the far end it is reflected back to the source. As might be expected, the magnetostriction effect is reversible. The mechanical deformation of the rod produces a change in the flux linking the grid coil that surrounds the left-hand end of the rod. The induced grid voltage varies the plate current, augmenting the reflected wave just as it returns. By this resonant periodic drive, the amplitude

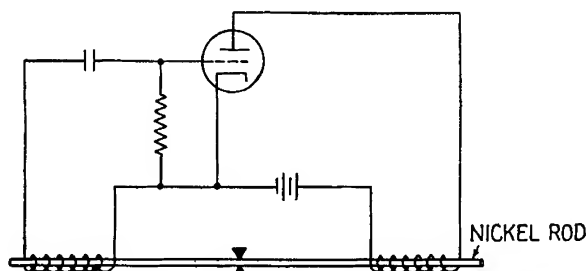


FIG. 9.1.—Basic diagram of magnetostriction oscillator.

of mechanical motion is built up. The rod is clamped or supported at the mechanical node, halfway between the ends of the bar. Thus the ends are free to move as coordinated pistons, producing a directive radiation of sound in the external medium. If within audible limits, the vibration of the rod is clearly heard. However, the device may be used also at frequencies well above the limits of audibility, producing a powerful supersonic wave. Preferably, the magnetic field of the plate coil is superposed upon a steady magnetic field of larger magnitude furnished by an adjacent bar magnet or horseshoe magnet. This improves sensitivity and also prevents the generation of two mechanical cycles for each electrical cycle, since the magnetostriction effect depends upon the magnitude of the applied field and not upon the magnetic polarity. The magnetostriction oscillator may *also* be regarded as a specialization of the Meissner circuit, Fig. 3.4, the vibrating rod replacing the electrical tank circuit.

Zero temperature coefficient may be obtained by using appro-

prate ferromagnetic alloys in place of pure nickel. The magnetostriction effect in iron and steel is relatively weak. Precision of frequency control may be made comparable with that of a precision crystal oscillator. In general, the magnetostriction oscillator is not directly in competition with the crystal oscillator, as their useful frequency ranges overlap only in a limited region, say from 50,000 to 100,000 cps. The magnetostriction oscillator is particularly convenient in the 1,000- to 50,000-cps region. Fixed plate and grid coils are effective over a considerable range of frequencies. Hence calibrated rods may be quickly interchanged for moderate alteration of frequency. Adjustment is effected by trimming the rod by means of a lathe, file, and emery cloth. The oscillator finds natural applications in underwater-sound technique, in the study of biological and chemical effects due to strong supersonic radiation, and in supersonic applications in general.

Like the quartz crystal, the magnetostriction rod is not limited in use to the stabilization of an oscillator. As a resonator it is a useful element in filter design. As an output load on a power amplifier it may drive a very heavy diaphragm, converting electrical energy into acoustic energy. In the supersonic range it is a useful microphone. Resonance that is too sharp may be damped whenever a special application demands a broader curve. Sufficient damping may often be obtained by replacing the solid rod with a nickel tube filled with lead. A wooden core, driven into a tube, is also effective.

Other types of mechanical stabilizers have been applied to self-driven vacuum-tube amplifiers but have not attained appreciable general use.

10. Beat-frequency Oscillator.—In theory, any one of the basic triode oscillators, previously described, is as usable in the audio-frequency range as in the radio-frequency range. In practice, certain limitations and difficulties are experienced. Very often an audio oscillator is a variable-frequency oscillator, covering an extensive range, say 20 to 20,000 cps. The unavoidable minimum capacitance of variable capacitors and associated wiring being taken into account, it is difficult to obtain a frequency range much larger than 1 to 2 by continuous variation of a single capacitor. Hence a wide-range oscillator of the simple basic type necessarily includes "band-changing" switches for exchanging coils, or adding extra inductance, or loading capacitance. Furthermore, all these audio-frequency inductors are heavy and relatively expensive.

Accordingly the elaborate circuit indicated in Fig. 10.1 may actually weigh less, cost less, and occupy less space than an equivalent single oscillator designed to cover a similar range. In addition, only a single dial is needed for calibration and control. This is a very considerable advantage in many applications. Within a single carrying case the device includes two oscillators and an amplifier. However, each oscillator is a radio-frequency generator; hence it does not require heavy metal-core inductors but instead uses light and inexpensive air-core coils. One of the oscillators operates at a fixed frequency, for example 1,000,000 cps. The other oscillator is variable, for example 1,000,000 to 1,020,000 cps. Since this variation is such a small percentage of the operating frequency, it is easily attainable by using a single dial control.

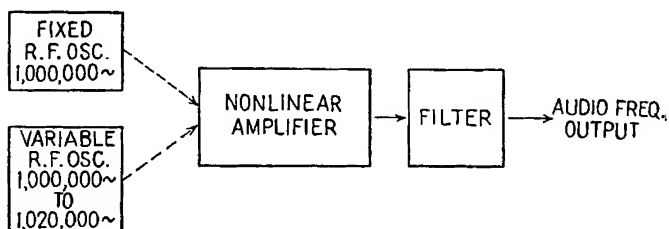


FIG. 10.1.—Beat-frequency oscillator.

To be specific, let us assume that this second oscillator has been set, for the moment, at a frequency of 1,012,000 cps. A simple linear addition of the two radio-frequency voltages would yield a wave of varying amplitude, the envelope rising and falling 12,000 times per second. However, despite the appearance of the 12,000-cycle-envelope variation (where plotted on paper or on an oscilloscope screen) there is no actual audio frequency present, and it is not advisable to refer to this envelope modulation as a “beat note.” In order to make a 12,000-cycle component actually available for driving a loudspeaker or other audio-frequency apparatus it is necessary to introduce rectification or some other deliberate non-linearity into the system. A “square-law” device would be sufficient. Hence the nonlinear amplifier indicated in Fig. 10.1, which makes the 12,000-cycle beat note available at the output terminals. In addition to the desired difference frequency, the nonlinear device will also produce the sum frequency, *viz.*, 2,012,000 cps, together with harmonics and combination tones of each of the high-frequency oscillations. However, these undesired components are generally far removed from the 12,000-cycle output frequency,

and hence they may be removed by a relatively simple and inexpensive filter.

By very careful shielding and decoupling, the two r-f oscillators may be made to operate independently of each other, right down to a "zero-beat" condition. With less attention to careful decoupling, the two oscillators tend to synchronize (to pull in step), once the difference frequency is reduced to a small margin, say 20 cps or less. In a modern well-designed beat-frequency oscillator, retention of such complete frequency control makes it possible to use the zero-beat condition as a quick and accurate check upon the adjustment of the "fixed-frequency" oscillator. If a built-in indicator such as a neon lamp fails to show a steady condition when the frequency control is turned to zero, then the fixed-frequency oscillator may be readjusted slightly until the error is corrected. After the apparatus has warmed up during the first few minutes of operation, occasional verification suffices.

In some older models, undesired synchronization prevents use of zero frequency as a check point on the frequency dial. In this case a similar test and adjustment are made at some other conveniently available frequency such as 60 or 1,000 cps. Sometimes a mechanical reed or vibrator is built in as a test device. In modern applications the beat-frequency oscillator sometimes is used to produce output voltages far above the audio-frequency range.

11. RC Oscillators.—*RC* oscillators are relatively new competitors of the beat-frequency oscillator, intended for the same general class of application, and possessing in even greater degree the advantages of light weight, small size, and low cost. As the name suggests, an *RC* oscillator is an electronic oscillator that employs only resistance and capacitance elements in its network, no inductors being required. Nonsinusoidal *RC* oscillators have been in use for many years. These are known as "relaxation oscillators" and are described under the general heading of Timing Circuits, Chap. XXIV. Sinusoidal oscillators of the *RC* type are a comparatively recent development.

The name *RC* oscillator refers, not to a single definite circuit, but rather to a wide variety of circuits with which comparable results may be obtained. As representative examples the "lag-line oscillator" and the "Wien bridge oscillator" will be described.

One type of lag-line oscillator is shown in Fig. 11.1. The vacuum tube, which may be a triode, tetrode, or pentode, is connected as a conventional resistance-coupled amplifier. On account

of the usual "seesaw" action of such a device, the plate potential falls as the grid potential rises—a phase shift of 180° if the grid voltage can be made sinusoidal. In order that this amplifier may be self-driven, some form of feedback is required. In order to be *regenerative*, the voltage fed back must arrive *in phase* with the grid voltage so that it may increase the amplitude of this voltage or at least maintain its level. Therefore the feedback circuit must function as a phase shifter, introducing an additional phase shift of 180° in the external route from plate back to grid. In addition, it must include some form of frequency-determining network, there being no other provision for determining the frequency at which oscillation will occur. These requirements are satisfied in a simple, straightforward fashion by the "lag line," shown at the right in Fig. 11.1. This is a "delay circuit," often placed in a vacuum-tube network in order to introduce a time lag into the operation. A

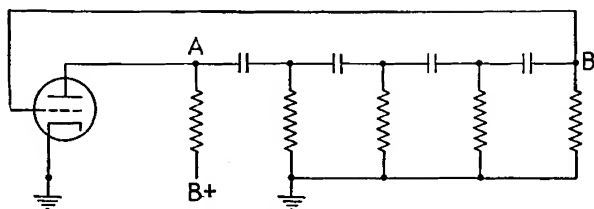


FIG. 11.1.—Phase-shift oscillator.

delayed and attenuated reproduction of the voltage applied at *A* arrives, after the intended interval, at point *B*. For some one particular frequency this time interval corresponds to a phase lag of 180° . At this selected frequency the over-all gain of the circuit is a maximum. This maximum gain may be adjusted so that oscillation just occurs. Theoretically, regenerative feedback is also emphasized at higher frequencies for which the phase lag is $3\pi/2$, $5\pi/2$, etc. However, at these frequencies the over-all gain is less, and the fundamental frequency prevails. An approximately sinusoidal waveform is attainable with moderate care in adjustment of gain by varying the supply voltage or by any convenient means.

The design of such an oscillator is discussed in the literature.¹ Several other types of ladder network, or artificial line, may be used as the lag line. The number of sections is not critical, and the frequency of oscillation is inversely proportional to the product RC .

¹ E. L. GINZTON and L. M. HOLLINGSWORTH, Phase-shift Oscillators, *Proc I.R.E.*, **29**, 43, 1941.

The frequency may be adjusted by varying one or more of the series capacitors. For maximum effect all of them are varied identically by a single shaft. Additional range may be secured by gang adjusting all the shunt resistors. While this is not quite so convenient as the wide-range single-dial control inherent in the beat-frequency oscillator, this feature could be obtained, if warranted, by gearing the *R* shaft and the *C* shaft together. Like the beat-frequency oscillator, *RC* oscillators in general find widest application in frequency ranges below 30,000 cps, where the elimination of inductors is particularly advantageous in saving weight, size, and cost.

In describing the six basic triode oscillator circuits, Sec. 3, reference is made to two common features that they all possess, *viz.*, (1) approximately 180° phase angle for grid and plate voltages; (2) provision for energy storage, with resultant “flywheel” effect.

The lag-line oscillator, typical of *RC* sinusoidal oscillators in general, also possesses these same two characteristics. The phase reversal is obtained by use of the lag line as a phase shifter. Energy storage is provided by the fact that energy is flowing on the “line” at all times and requires a finite time for transmission. The line capacitors are charged and discharged in succession as the wave is handed on from one section to the next.

12. Wien Bridge Type of *RC* Oscillator.—The simple requirements of an oscillator circuit also can be met by a number of other

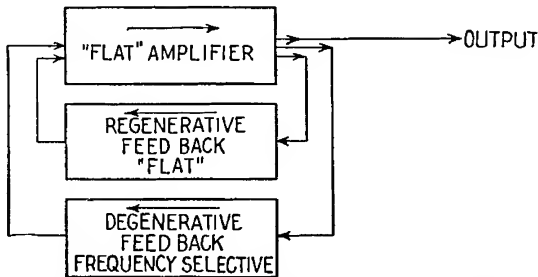


FIG. 12.1.—Block diagram of Wien bridge oscillator.

phase-shifting networks constructed of resistors and capacitors only. The Wien bridge oscillator is indicated in block diagram form in Fig. 12.1. The *RC* amplifier is conventional, having a constant amplification, at least over the range of frequencies to be derived from the oscillator. The regenerative feedback is also “flat” over a similar frequency range. Hence this return channel

tends to promote oscillation at any or all frequencies over a wide band. However, the *regenerative* feedback is directly in competition with a *degenerative* feedback, an additional return channel through which a part of the output is available for opposing any grid-voltage variation. In general, the degeneration feedback is stronger and therefore successfully prevents oscillation, but at one particular frequency the output voltage from the degenerative feedback channel is zero. At this frequency the regenerative voltage wins by default, and the oscillation occurs.

Evidently the key component of the oscillator is the degenerative feedback channel. It should transmit alternating voltages freely, in general, but should sharply discriminate against transmission of a single definite frequency. This description precisely

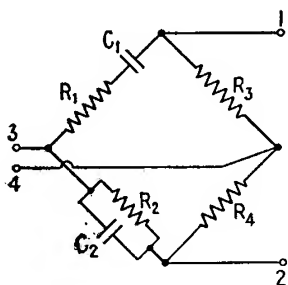


FIG. 12.2.—Wien bridge.

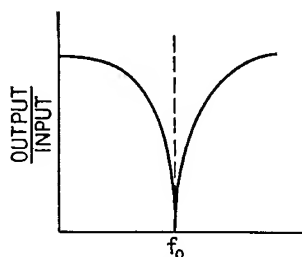


FIG. 12.3.—Frequency response of Wien bridge.

fits the behavior of a “frequency bridge”—a bridge that has a null balance at a definite frequency and is thrown out of balance by a minor deviation of frequency. Various modifications of the basic Wheatstone bridge possess this property. In ordinary engineering practice they have been used for the measurement of resistance and reactance at a known frequency or for the determination of an unknown frequency in terms of known circuit constants. For prospective use as a circuit component in an *RC* oscillator several of these bridges are eliminated by the fact that they include one or more inductors as bridge elements. Others are undesirable because frequency discrimination is not sufficiently sharp.

The Wien bridge, illustrated in Fig. 12.2, fits the specifications particularly well. In general, application of a voltage at the input terminals 1-2 results in the production of an attenuated output voltage at 3-4. However, a sharply defined null, Fig. 12.3, is obtained at a frequency f_0 such that

$$f_o = \frac{1}{2\pi R_1 C_1}$$

provided that the following conditions are satisfied in the construction of the bridge,

$$\begin{aligned} R_1 &= R_2 \\ C_1 &= C_2 \\ R_3 &= 2R_4 \end{aligned}$$

Hence the frequency f_o at which oscillation will occur may be varied by simultaneous variation of the equal capacitors C_1 and C_2 , gang operating these variable capacitors by a common shaft. For any given design of variable capacitor, necessarily involving an irreducible minimum capacitance, much wider range of frequency adjust-

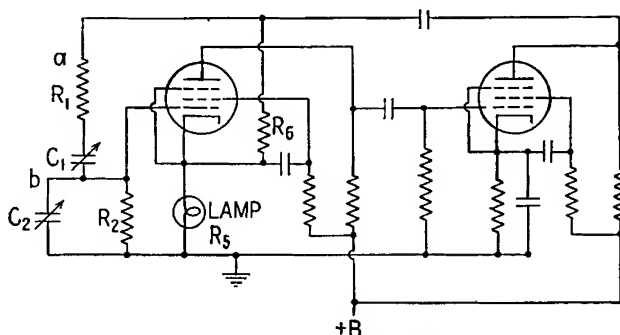


FIG. 12.4.—Resistance-stabilized RC oscillator.

ment may be obtained in the Wien bridge oscillator, and in similar RC oscillators, than is obtainable in any of the simple LC oscillators. The inverse first-power dependence upon C_1 is an improvement over the inverse square-root term that governs the frequency of the conventional tank circuit. Additional frequency range may be provided, if required, by simultaneous variation of R_1 and R_2 .

Figure 12.4 is a type diagram of a very similar oscillator¹ that is commercially available.² In this type the *degenerative* feedback is provided by resistances R_5 and R_6 and is substantially independent of frequency. Use of a small incandescent-lamp filament as resistor R_5 provides automatic increase of degenerative feedback as the amplitude of the signal increases. This limits the oscillation

¹ F. E. TERMAN, R. R. BUSS, W. R. HEWLETT, and F. C. CAHILL, *Proc. I.R.E.*, **27**, 649, 1939.

² Hewlett-Packard Company, Series 200 audio oscillators.

amplitude to a convenient small-signal value that ensures approximate linearity of operation and consequent sinusoidal waveform. The *regenerative* feedback is provided by the $R_1C_1R_2C_2$ network, which is frequency selective. With $R_1 = R_2$ and $C_1 = C_2$, correct phase angle and maximum feedback voltage amplitude occur at the frequency

$$f_o = \frac{1}{2\pi R_1 C_1}$$

at which oscillations occur.

In comparison with the conventional Wien bridge oscillator, it should be noted that resistors R_3 and R_4 , Fig. 12.2, have been suppressed and the frequency-dependence characteristics of the regenerative and degenerative feedback paths have been interchanged. Strictly, therefore, Fig. 12.4 is not a Wien bridge oscillator, though the designation has been used occasionally.

13. Negative-resistance Oscillators.—In general terms, all oscillators may be regarded as negative-resistance devices. Upon

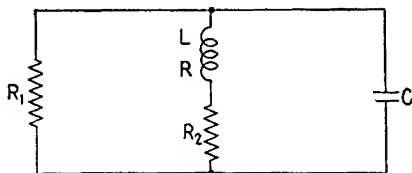


FIG. 13.1.—General principle of the negative-resistance oscillator.

opening the circuit at any arbitrary point and considering the input admittance offered to a current of appropriate frequency, a finite current is built up without requirement of an external driving voltage. Such a concept already has been employed in considering the tuned-grid oscillator, for example.

Traditionally, however, the phrase “negative-resistance oscillator” is used in a more restricted sense. It refers to an LC tank circuit in which continuous sinusoidal oscillations are built up and maintained by a simple two-terminal attachment to the power circuit. The entire power circuit, however elaborate, is equivalent to a series or shunt negative variational resistance. In Fig. 13.1, for example, oscillations are possible provided that either R_1 or R_2 offers a negative resistance of sufficient magnitude.

The “singing arc” was an early device of this character, still useful as an example, since it illustrates the general principle without employing complex circuits. In Fig. 13.2, the current

from the d-c generator is maintained at an approximately constant value by the heavy iron-core choke coils connected in series in the supply mains. The arc is an inherently unstable load. Current fluctuations occur constantly as the arc stream wanders around the white-hot crater. Moreover, any momentary increase in current raises the crater temperature and results in more copious thermal emission of electrons. Because of the great increase in available charge carriers, the larger current is maintained with a *lower* voltage drop across the arc. Hence the current-voltage characteristic of the arc has a *downward* slope illustrated in Fig. 13.3. An increment in current corresponds to a decrement in voltage, and vice versa. Hence the variational resistance of the circuit element is negative. In the right-hand mesh of the network, this negative resistance is included in series with L and C and in series with the

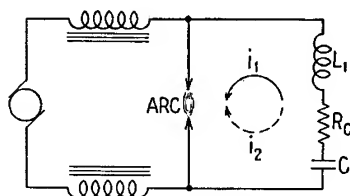


FIG. 13.2.—“Singing arc.”

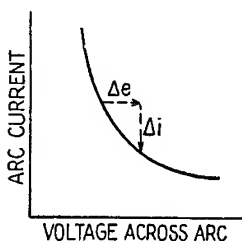


FIG. 13.3.—Negative variational resistance of arc.

normal positive resistance R of the coil and load (R_2 negative in Fig. 13.1).

On the average, the capacitor C will assume a charge consistent with the average voltage maintained across the arc, say 100 volts. Soon after this general equilibrium is established, a momentary fluctuation increases the arc current. The generator cannot supply immediately the extra demand on account of the large choke coils in the supply circuit. Instead, the line current increases only by a small amount, just sufficient to alter the flux through the reactors, producing a resultant drop of voltage at the arc. The drop in arc voltage allows capacitor C to discharge partly, thus producing the circulating current i , which in turn supplies most of the increase in arc current.

The electrical inertia of inductor L_1 causes current i_1 to persist, even after the voltage across capacitor C has dropped to the level required by the lowering of arc voltage. Hence the capacitor C is *overdischarged* when current i_1 finally ceases. A charging current

i_2 immediately replaces i_1 . Being in the opposite direction around the loop, current i_2 is drawn from the generator at the expense of the current in the arc stream, the line current remaining nearly constant. This results in an increase of arc voltage, aiding and abetting the capacitor-charging process. Electrical inertia overdoes this process also, and a new cycle of discharge begins. Amplitude of oscillation increases until the negative variational resistance no longer exceeds the real positive resistance R . The instability of the arc causes it to act as a valve, routing the generator current alternately into the capacitor and into the arc, at a frequency determined primarily by the LC product of the tank circuit.

In the Poulsen arc oscillator, once widely employed as a long-wave radio-frequency generator, the instability of the arc was emphasized deliberately, and the changes of voltage and current were speeded up so that the arc could follow the cycle of a sufficiently slow radio-frequency oscillation. To this end, various expedients were adopted, such as operation of the arc in hydrocarbon vapor, use of special copper and carbon electrodes, water cooling, and operation of the arc in an intense magnetic field. Such devices were finally made obsolete by the development of high-power vacuum tubes, using the conventional triode circuits previously discussed.

The next step in the application of the negative-resistance type of oscillator came with the introduction of the *dynatron*, about 1918. Making use of secondary emission in an ordinary triode or tetrode, an electronic network equivalent to a simple negative resistor is bridged across an ordinary LC tank circuit (R_1 negative in Fig. 13.1). The presence of the negative resistance may be deduced as follows:

In Fig. 13.4 let the grid be held at a fixed positive voltage, say +50 volts, while the plate potential e_p is increased by steps from zero to 50 volts. The corresponding plate current i_p is measured by an ammeter. For $e_p = 0$, practically no electrons arrive at the plate, and the plate current is practically zero. For small positive values of e_p , some of the electrons that happen to pass between the grid wires are drawn to the plate and are captured, registering a positive current in passing through the ammeter. The increase of current with increasing voltage is normal at first; but, with increasing impact velocity, secondary emission begins. Such secondary emission from the plate undoubtedly occurs copiously in most vacuum-tube applications, but normally it produces no

observable external effect. In normal operation of a triode amplifier, for example, the plate is the most positive electrode in the tube, and hence secondary electrons knocked from the plate return to it without contributing to the external current. In the dynatron, however, the secondary electrons ejected from the plate are drawn away to the more positive grid, thus resulting in a net loss of plate current. Thus the i_p curve rises at a smaller and smaller rate,

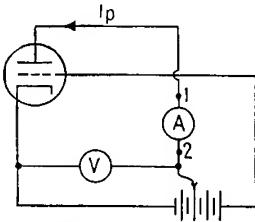


FIG. 13.4.—Measurement of secondary-emission effects.

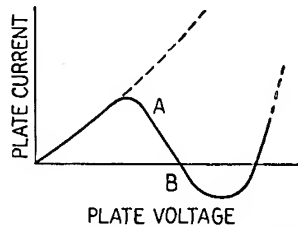


FIG. 13.5.—Dynatron characteristics.

finally becomes horizontal, and then starts downward. In this important negative-slope region the secondary emission increases more rapidly than the primary current to the plate. Eventually the net current becomes zero and then reverses, the number of secondary electrons exceeding the number of primary electrons. The negative-slope region is finally terminated when the plate approaches the grid potential, thus weakening the field that is needed in order to sweep the secondary electrons away. (This field may be compared to a draft of air sweeping water-vapor molecules away from the surface of a liquid and allowing further evaporation to take place.)

The existence of a negative slope region, AB , Fig. 13.5, implies that the entire circuit presents a negative variational resistance when seen from the ammeter terminals 1-2, Fig. 13.4. Hence on disconnecting the ammeter, one may substitute a conventional LC tank circuit, thus obtaining the dynatron oscillator, Fig. 13.6. The simplicity of the two-terminal connection to the oscillating circuit is sometimes advantageous. For example, one can convert an ordinary wavemeter (a calibrated LC circuit) into a self-excited signal generator by merely mounting a small vacuum tube and battery power supply on the binding

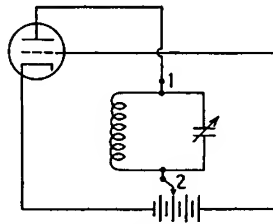


FIG. 13.6.—Dynatron oscillator.

posts of the wavemeter. The slight change of calibration can be readily determined.

However, the dynatron has limitations as well as advantages. Any device dependent upon secondary emission is likely to have somewhat variable operating characteristics. This is particularly true of the triode dynatron for we are abusing the grid by subjecting it to a steady positive voltage. Overheating of the grid wires is likely to liberate gas from the metal, even if no actual melting takes place. The life of the tube is likely to be shorter than in conventional operation, and operation may be erratic if the gas pressure changes. In any event the power output is limited to small values

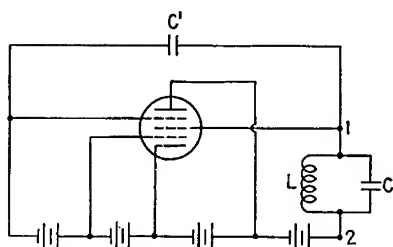


FIG. 13.7.—Transitron oscillator.

by the necessity of avoiding excessive power-supply voltages. Some improvement is secured by using a tetrode, as the screen grid is better adapted for operation at a steady positive voltage. However, if a multielectrode tube is to be used at all, it is generally better to use the transitron circuit described below.

The transitron connection is a more elaborate method for obtaining electronically a two-terminal negative resistance to be shunted across an LC tank circuit. In Fig. 13.7 capacitor C' may be regarded as the key element of the circuit. The capacitance of C' is not critical, but it must be large enough to ensure that the suppressor grid will follow, without appreciable amplitude decrease or time lag, the variations in voltage imposed by the screen grid and its attached LC circuit. The average values of suppressor and screen potentials are very different, as is indicated by the power-supply connections, but their *variations* from the average are intended to be identical. In the absence of C' a decrease in screen voltage would result normally in a corresponding decrease of screen current. However, with C' in circuit, the coincident drop in suppressor-grid potential turns back toward the screen a great number of electrons that would otherwise have continued to the plate. This results in a net *increase* of screen current. Conversely, a simultaneous increase in screen and suppressor potentials produces a net *decrease* of screen current. Again substituting an LC tank circuit in place of an ammeter at terminals 1-2 the device becomes a self-excited oscillator. More elaborate than the simple dynatron,

it retains the advantage of a two-terminal attachment to the oscillating circuit and yet is not dependent upon secondary emission. As it does not involve maltreatment of tube elements, the transitron may be designed to produce a reasonable power output without curtailment of tube life.

Like the dynatron, the *magnetron* has been in use for a long time. Originally it was a simple diode, having a cylindrical anode, with a hot-wire emitter along the axis of the cylinder. Outside the glass envelope, magnets or coils were used in order to produce a constant and uniform axial magnetic field. Electrons emitted from the hot wire traveled in uniform spiral arcs, Fig. 13.8, experiencing a transverse deflecting force dependent upon the strength of the steady magnetic field. On gradually increasing the magnetic-field intensity the plate current showed no change until, suddenly, at a critical value of H , the current dropped to zero, Fig. 13.9. The increasing magnetic field had decreased the radii of the spiral arcs until they abruptly failed to make contact with the anode. Aside from minor deviations due to accidental differences in the emission velocity, the spiral orbits should be identical and should all break contact at the same critical value of magnetic-field intensity. Among other

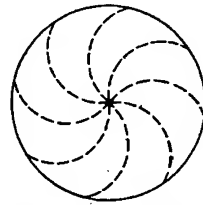


FIG. 13.8.—Simple magnetron.

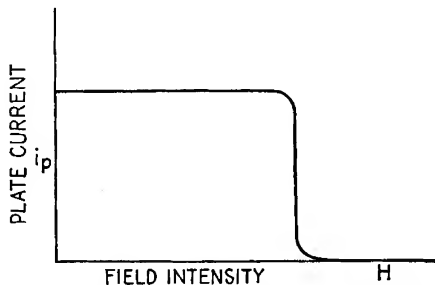


FIG. 13.9.—Characteristic curve of simple magnetron.

applications the early magnetron was proposed as a switching device for conveniently interrupting high-voltage d-c transmission-line currents.

The next step in the development of the magnetron came when the cylindrical anode was split into two separate semicylinders. The two half sections were maintained at the same average potential by means of a common power supply, Fig. 13.10. In the quiescent

state the orbits of the electrons were practically unaffected by the splitting of the anode. However, any transient oscillation in the LC circuit, however slight, would cause the two anodes to vary from the quiescent state in opposite directions, one anode becoming slightly more positive, the other less positive by an equal amount. Because of the spiral orbits, electrons initially accelerated toward the more positive plate would not necessarily be collected by that plate. By experiment and by theory it was shown that an increase of voltage on the upper plate actually reduced the net current to

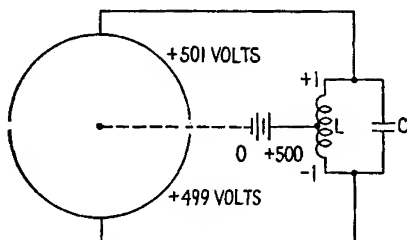


FIG. 13.10.—Split-anode magnetron.

that plate and increased the current to the lower plate. If accompanied by a simultaneous lowering of the potential of the lower plate, the inequality of current flow was augmented. In effect, two negative resistors were shunted across the halves of the tank coil L , operating in push-pull cooperation to build up the oscillation in the tank. Further development of the magnetron has led to the multiple-plate super-high-frequency device.¹ This has been employed widely in high-power “microwave” applications.

¹ Described in *Proc. I.R.E.*, **32**, 136, 1944.

CHAPTER XVI

GAS-FILLED TUBES

1. Electrical Discharges in Gases.¹—When the voltage applied to two cold electrodes in a gas is increased from zero, only a very small current (of the order of microamperes) flows until a critical voltage is reached. This critical voltage, shown at point A, Fig. 1.1, is called the sparking voltage. At the sparking voltage the

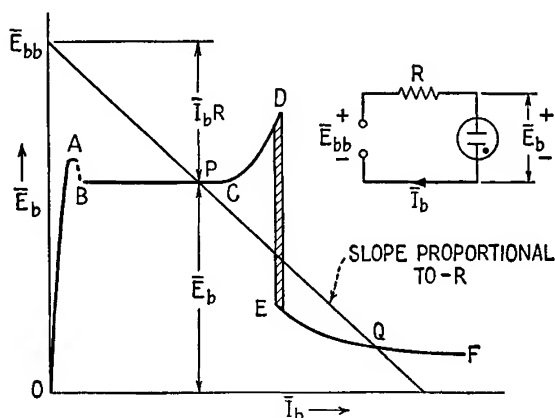


FIG. 1.1.—Gas-discharge characteristics.

nearly nonconducting gas becomes a good conductor, and the current that flows is determined by the series resistance, the discharge characteristic, and the applied voltage, Fig. 1.1. The passage of the initial spark causes the formation of large numbers of positive ions and of electrons in the path of the discharge. Actually, less than 0.1 per cent of the atoms are ionized, but the number of atoms per cubic centimeter, even at low pressures, is quite large. The electrons are accelerated toward the anode by the applied field, and when they gain sufficient energy they cause additional ionization by knocking electrons out of neutral atoms with which they collide. The positive ion resulting when a neutral atom loses an electron is

¹J. D. COBINE, "Gaseous Conductors," Chap. XI, McGraw-Hill Book Company, Inc., 1941.

massive compared with the electron, so that it moves relatively slowly toward the cathode. However, a few of the positive ions gain enough energy to cause the emission of electrons when they strike the cathode. This emission by positive-ion bombardment maintains the discharge.

From *B* to *C*, Fig. 1.1, the discharge is characterized by a voltage drop that is practically independent of the current and is usually of the order of 100 volts, varying from about 70 to 400 volts depending on the gas and the electrodes. This type of discharge is known as a *normal glow*. In the glow discharge the gas between the electrodes is luminous, the greatest intensity of glow being near the cathode in what are known as the cathode-glow and the negative-glow regions. Tests indicate that for a normal glow the current density at the surface of the cathode is constant. As the current is increased, the area of the cathode covered by the glow increases. At the point *C*, Fig. 1.1, the cathode area is completely covered by the cathode glow, and a further increase in current is accompanied by an increase in discharge voltage. What is known as an abnormal glow is characterized by the rising portion *CD* of the curve. If the current in the vicinity of *D*, usually of the order of 100 ma, is increased and if the cathode is cold, a sudden transition occurs to a low-voltage high-current type of discharge called an *arc*. In the arc at low gas pressures the conducting path glows similarly to the glow discharge, but the light is much more intense owing to the greater current of the arc. The arc voltage drop is quite low compared with that for the glow, being usually from 10 to 15 volts. The arc voltage drop decreases as the current increases, becoming nearly constant at large currents. If the arc path is constricted by girds, metal shields, etc., the arc voltage may increase with current owing to the loss of ions to these surfaces. The low voltage drop of the arc is due to a more efficient mechanism of electron emission from the cathode than that of the glow. In tubes in which arcs are employed, the cathode is usually incandescent, and thermionic emission is a much more efficient process than the positive-ion-bombardment process of the glow discharge.

The voltage distribution between cathode and anode for the gas discharge is compared in Fig. 1.2 with that in vacuum, with and without the space-charge distortion resulting from electron emission. With plane-parallel electrodes in vacuum and with no electron emission, the potential varies uniformly. If the cathode is emitting electrons so that the current is limited by space charge,

the potential follows the distribution indicated, resulting in the three-halves-power law for plate current, Chap. X, Sec. 4. However, when sufficient gas is present for a glow or an arc discharge to occur, the greater part of the voltage drop occurs very near the cathode in what is known as the cathode-drop region. In the cathode-drop region there is a high concentration of positive ions, which neutralizes the negative space charge due to the emitted electrons. It is in this region that the positive ions gain sufficient energy to cause the emission of electrons when they strike the cathode. In the luminous region, called the plasma, extending

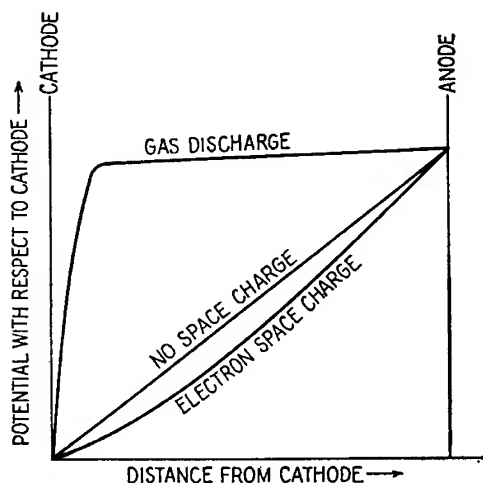


FIG. 1.2.—Voltage distribution in gas discharge and in vacuum (plane-parallel electrodes).

from the cathode-drop region to the anode, equal concentrations of positive ions and electrons exist. This provides a path of high conductivity so that a very low electric field strength can maintain the current.

2. Glow Tube.—The normal glow discharge, region BC , Fig. 1.1, may be used as a voltage regulator by placing the load circuit directly across the tube. In this application variations in supply voltage \bar{E}_{bb} are absorbed by the series resistance R as the current through the tube varies along BC ; or if the load resistance changes, the current through the glow tube so changes.

The OD3/VR150 voltage-regulator tube is sketched in Fig. 2.1. VR stands for voltage regulator. The number 150 is the approximate operating plate voltage in the flat region of curve a , Fig. 2.2.

The I.R.E. tube symbol indicates the anode (for regulator operation) by a straight line, as for a triode, and indicates the circular cathode by a small circle. Actually, the cathode is the larger of the two electrodes, and to emphasize this fact an additional symbol is used

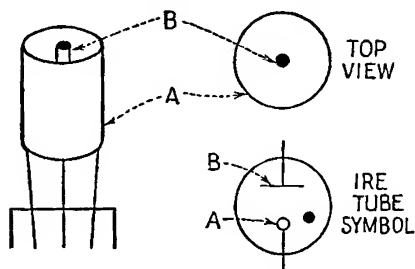


FIG. 2.1.—OD3/VR150 glow-discharge tube.

in Figs. 2.2 and 2.3. Elsewhere the conventional I.R.E. symbol is used to indicate a voltage-regulator tube.

A glow-discharge tube having electrodes of widely different areas in a gas at a pressure of the order of 2 cm of mercury can be used to rectify alternating current. When the large electrode is negative, the voltage drop across the tube is relatively small and is independent of the current, Fig. 2.2, curve *a*. When the small

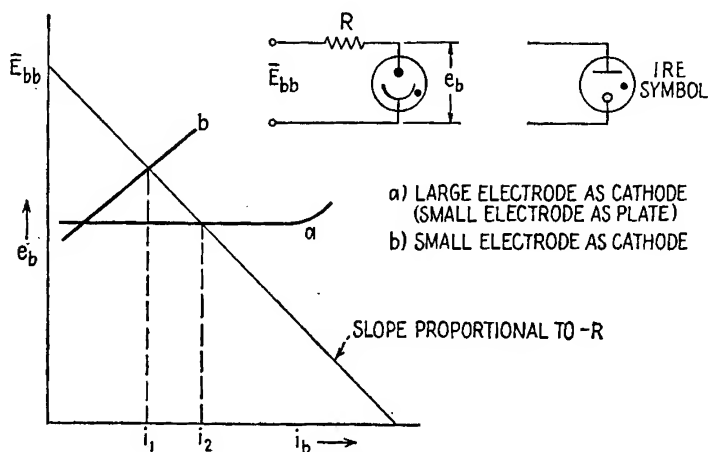


FIG. 2.2.—Glow-tube characteristics.

electrode is negative, the condition of abnormal glow is reached at a low current and the voltage drop across the tube increases rapidly with current, Fig. 2.2, curve *b*. Over a wide range of supply voltages, the current that flows when the small electrode is negative,

such as i_1 , Fig. 2.2, is less than the current, such as i_2 , that flows when the large electrode is negative, and the alternating-current wave is unsymmetrical and has an average value \bar{I} , Fig. 2.3. This type of rectifier is characterized by a relatively high voltage drop, of the order of 100 volts, and a small current so that it is inefficient. Since no filament power is required, it is useful in some low-power applications. The direction of sweep and the polarity of a cathode-ray oscillograph may be determined by observing the characteristically unsymmetrical voltage wave of a glow-discharge rectifier tube operating on alternating voltage.

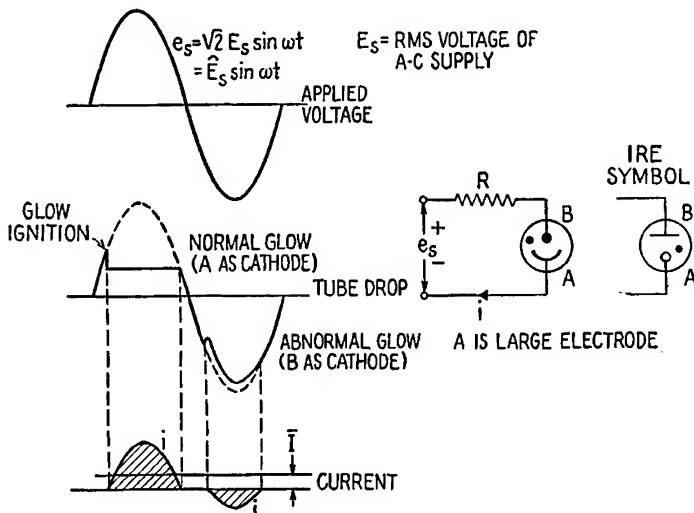


FIG. 2.3.—Glow-tube rectification.

A third, or control, electrode is sometimes added so that these tubes can be used for low-power relay applications.

3. Gaseous Thermionic Diode.—A very small amount of chemically inert gas admitted into a thermionic diode will produce an increase in the plate current for plate voltages less than that corresponding to current saturation in vacuum, Fig. 3.1. This increase in plate current is due to some of the gas atoms becoming ionized. Positive ions, formed when atoms are ionized by electrons accelerated from the cathode, move slowly toward the cathode. The velocity of the positive ions, because of their relatively large mass, is several hundred times slower than that of electrons in the same electric field. Because of the large difference in velocities, one positive ion can neutralize the effect of many passing electrons.

Thus the positive ions partly neutralize the negative space charge produced by electrons from the cathode, and a given plate voltage causes a larger electron current. This effect is confined to the range of potentials in which current is limited by space charge, lying between the ionization potential of the gas and the plate potential at which the saturation emission of the filament is reached.

The positive ions striking an oxide cathode may cause its gradual destruction by sputtering, *i.e.*, knocking off atoms composing the coating material.

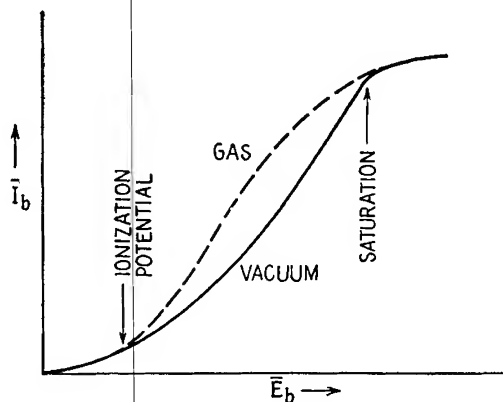


FIG. 3.1.—Effect of trace of gas in a thermionic diode.

4. Arc Rectifiers.—When sufficient gas is present in a thermionic diode, an arc is established between the electrodes. The thermionic cathode supplies all the necessary electrons so that only a low voltage is required to start and maintain the discharge. The arc discharge starts with a voltage slightly greater than that necessary to sustain the arc. The characteristic volt-ampere curve of the thermionic arc diode is compared with that of the thermionic vacuum diode in Fig. 4.1. The thermionic arc diode is a very efficient rectifier of alternating current owing to the low voltage drop in the tube. Rectification occurs because an arc is rapidly established at a low voltage when the cathode is made negative relative to the plate, while a high voltage is necessary to cause the emission of electrons from the plate and to start a discharge when the cold plate is made negative.

The *tungar*, or *rectigon*, is an arc tube filled with argon at a pressure of about 2 cm of mercury. It has a tungsten-filament cathode. The arc discharge formed between the hot cathode and

the cold plate has a voltage drop of the order of 10 volts, the drop being nearly independent of current. The tube can be operated at several amperes. The relatively high gas pressure permits the operation of the filament at a temperature higher than in vacuum. Therefore, a greater emission current is possible without danger of evaporation of the cathode material. Because of the relatively high gas pressure, a glow discharge can form with its cathode on the cold plate during the half cycle of alternating voltage when this electrode is negative if the inverse voltage is greater than about 200 volts. The formation of this glow is undesirable since there is the possibility of the formation of an arc cathode spot on the plate, with the result that the device loses its unidirectional characteristic. For this reason the tungar tube is used only with a-c sources of the

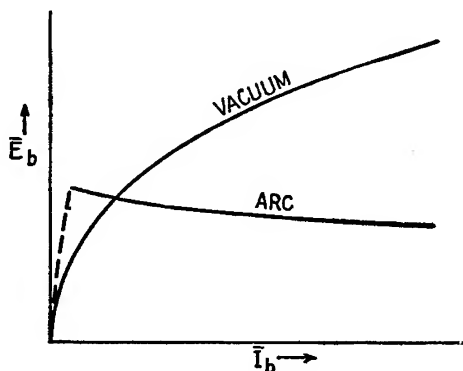


FIG. 4.1.—Characteristics of thermionic arc diode and thermionic vacuum diode.

order of 110 volts and is especially adapted for storage-battery charging.

The *phanotron* is a thermionic diode type of rectifier tube using an arc discharge in mercury vapor or in argon. The cathode used to supply the relatively large current of the arc may be an oxide-coated ribbon or an indirectly heated oxide-coated cylinder arrangement, Fig. 4.2. The surfaces of the vanes and the inner surface of the outer cylinder are brought up to the emitting temperature as in an oven. A very large electron current may be obtained from this type of cathode, which can function only because of the presence of gas in the tube. Electrons emitted within the cavities between the vanes make many collisions with gas atoms and thus can migrate to the open end, which would be impossible in the case of a vacuum. The arc drop is relatively constant as the current is varied and is of the order of 10 volts for mercury and 15 volts for argon. Since

the gas pressure of the phanotron is low, it can safely withstand a much higher inverse voltage than is possible with the tungar rectifier. When mercury vapor fills the tube, the vapor pressure is determined by the temperature of the coldest spot where mercury may condense. Thus the pressure in these tubes varies widely, being very low when the tube is cold and increasing to about 0.05 mm of mercury when the tube is heated to the normal operating temperature. Tubes filled with inert gases naturally do not have this wide pressure variation.

Precautions.—To avoid destruction of the cathode the tube drop of all thermionic arc tubes must not exceed 22 volts. A higher tube drop than this will result in destruction of the oxide coating by

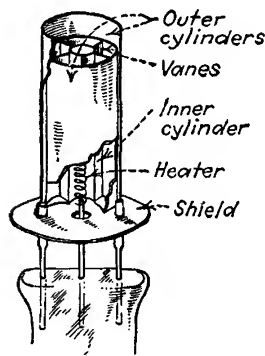


FIG. 4.2.—Indirectly heated cathode of arc-rectifier tube.

positive-ion bombardment. The tube drop is kept low by allowing the cathode to reach its normal operating temperature before applying voltage to the plate of the tube. Plate voltage is withheld for sufficient time for the cathode to reach its normal operating temperature and for the mercury to be vaporized to the correct pressure. Usually 5 min is sufficient heating time for a large tube; but if the tube has not been in a vertical position, mercury may be deposited on the elements and 15 min must be allowed for heating. Gas-filled rectifier tubes should be shielded from high-intensity high-frequency fields that could ionize

the gas during the normally nonconducting period and result in an arc-back.

Occasionally it is desirable to increase the power of a rectifier unit by operating several anodes in parallel. This can be done with thermionic vacuum rectifiers because of their rising volt-ampere characteristic. Since the normal volt-ampere characteristic of arc tubes has a negative slope, the anodes will not operate stably in parallel unless some resistance or a suitable interphase reactor is placed in each anode lead. The negative slope of the arc volt-ampere characteristic is equivalent to a negative resistance; therefore, sufficient resistance must be placed in series with the arc so that the net resistance is positive over the operating range of current. Usually resistors that have a drop of about 25 volts at rated current may be used in each anode lead.

5. Gaseous Thermionic Triode, or Thyatron.—The *thyatron* is a grid-controlled thermionic arc-type rectifier. The tube is similar in construction to the phanotron, except that the grid is added, Fig. 5.1. The object of the grid is to prevent the tube from conducting unless the grid potential is made more positive than a critical value. Figure 5.2 shows dependence of plate current upon grid voltage for the vacuum and the arc tubes. At a critical value of grid voltage the grid permits an arc to be established, and the current increases suddenly to the value determined by the applied voltage and the circuit resistance. Once the arc is started, the grid has no further control of current.

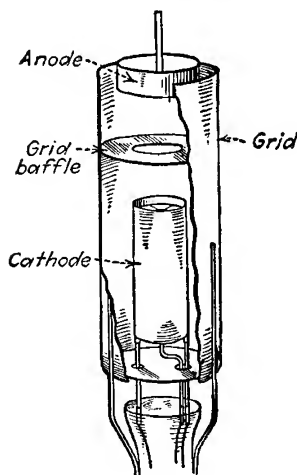


FIG. 5.1.—Electrode structure of negative-grid thyatron (GE type FG-57).

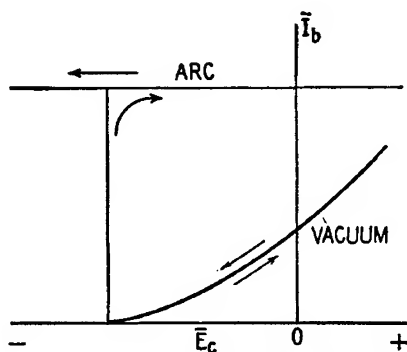


FIG. 5.2.—Effect of grid voltage on plate current of arc and vacuum diodes.

The relation between the critical grid voltage at which an arc starts and the anode voltage is shown on the left-hand side of Fig. 5.3. This curve is called usually the *grid-control characteristic*. Since the grid-control characteristic is nearly linear, its slope is sometimes spoken of as the *control ratio* of the tube. However, the slope varies from tube to tube and with the operating temperature so that it is not always a reliable reference. Figure 5.3 shows that as the anode voltage is increased a higher negative voltage is required on the grid to prevent conduction. Once an arc is formed between cathode and anode, the grid cannot stop the discharge even if it is made more negative than the critical value, Fig. 5.2. In order to stop the discharge, the plate voltage must be reduced to

zero for an instant. When this occurs the discharge path is deionized and a negative grid then can resume control. The deionization time varies from 100 to 1,000 microseconds depending on the construction of the tube. When an alternating voltage is applied to the plate, the grid can resume control at the end of each positive half cycle when the plate voltage becomes negative. The right-hand side of Fig. 5.3 shows the principle of operation when an alternating voltage is placed on the anode. The *grid-control locus*, showing the value of critical grid voltage at which conduction

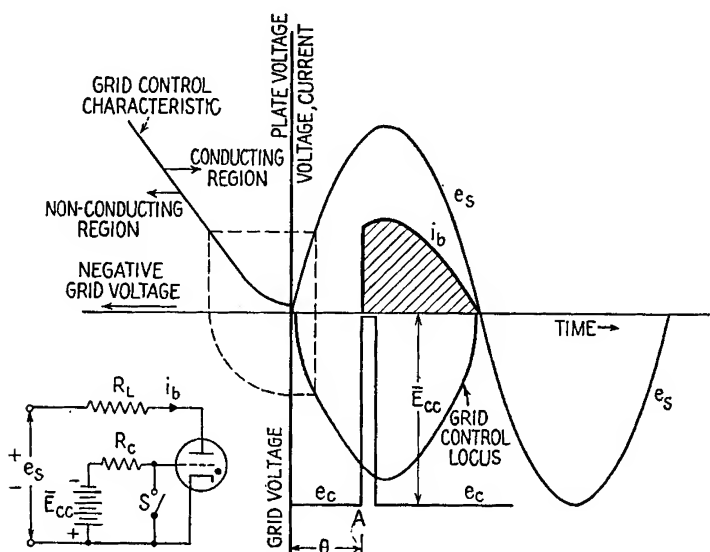


FIG. 5.3.—Control characteristics of a thyatron.

begins at any instant of a positive half cycle, is derived from the sinusoidal anode voltage wave e_s and the grid-control characteristic shown by the dotted lines, Fig. 5.3. If the grid has a constant negative bias E_{cc} , as by a battery, the tube can be made to conduct by impressing a positive pulse on the grid. Such a pulse is shown at A, obtained by momentarily closing switch S. If the load is a pure resistance, the load current that flows for the remainder of the cycle is shown by the heavy curve bounding the shaded area. By varying the instant at which conduction occurs, *i.e.*, by varying the firing angle θ , the average value of the rectified current for a single anode rectifier controlled as in Fig. 5.3 decreases as the firing angle θ increases, Fig. 5.4.

Figure 5.5a shows a conventional a-c control circuit for a thyatron rectifier. The autotransformer T , resistor R , and capacitor C form a phase-shifting circuit that does not require a d-c bias voltage on the grid. The phase angle θ , between the anode supply voltage e_s and the intersection of the e_{cc} curve and the grid-control locus,

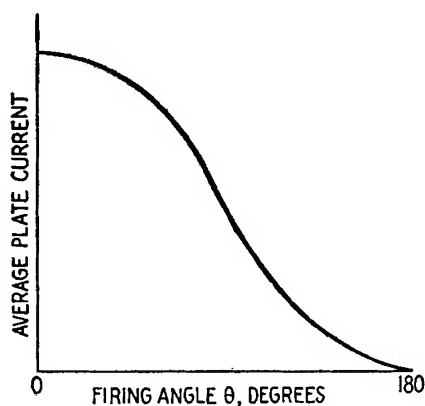


FIG. 5.4.—Control of average current of a thyatron.

Fig. 5.5, may be varied by changing the value of resistor R . The rectified current i , flowing through the load resistor R_L , may be varied from a maximum value to zero as the angle θ is varied from 0 to 180° by the phase-shifting circuit. It should be noted that if the a-c voltage on the grid is shifted beyond the 180° point relative

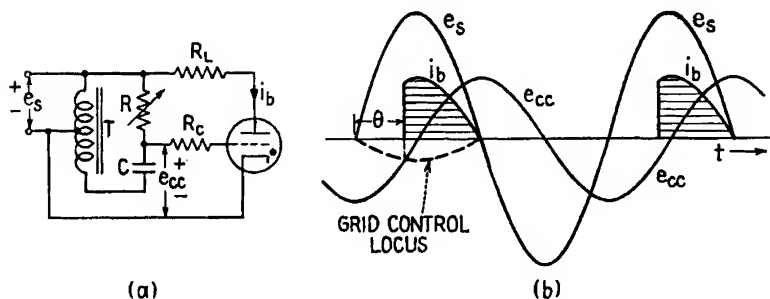


FIG. 5.5.—A-c control of a thyatron.

to the anode voltage, as by an induction phase shifter, the grid voltage will be positive but decreasing when the anode voltage becomes positive so that a maximum current will flow. This is true because even though the grid becomes negative it cannot stop the arc as long as the anode is positive. Other applications of the

thyatron as a current-controlling device are described in Chap. XXIV.

Extremely sensitive thyratrons are made by screening the control grid from both cathode and anode. Such screen-grid thyratrons can control considerable power with grid currents of the order of microamperes.

Precaution. Since the grid of a thyration is a relatively large metal surface similar to the anode, a high resistance must always be placed in series with the grid and the source of grid voltage as R_c of Fig. 5.5. This is essential in order to keep the grid-current less than the peak value specified by the manufacturer. Most tubes function satisfactorily if the grid resistance is of the order of 1,000 ohms per peak grid-supply volt when the grid is positive.

6. Current Ratings of Arc Tubes.—Arc tubes have the following current ratings, which must be carefully observed to prolong the life of the tube.

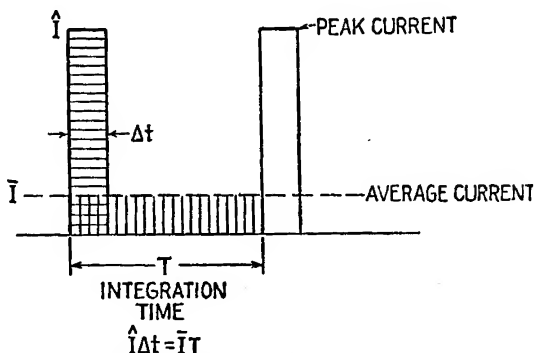


FIG. 6.1.—Graphical representation of average current, peak current, and integration time.

Average current, which is indicated by a d-c ammeter for a reasonably rapid cycle of operation. This current limit is set by the ability of the tube to dissipate heat, and of the glass seals to withstand thermal stresses.

Peak current, which is the greatest instantaneous current that can be supplied by the cathode without damage, Fig. 6.1. This is the saturation emission of the cathode under normal operating conditions. The peak current averaged over a period of time called the *integration time* must not exceed the average current rating. Arc tubes are not recommended for charging a capacitor. If they are used for this purpose, the capacitor should be considered as a

short circuit, which it is when voltage is suddenly supplied to an uncharged capacitor; therefore a series resistor must be placed in the circuit to limit the tube current to the peak current rating.

Surge current, which is the greatest current that can be passed accidentally once in the life of the tube. In the design of circuits sufficient reactance or resistance must be included to limit the current to the surge-current value should the load be short-circuited. Fuses should be provided to ensure that this value is never reached.

Grid current, which must be limited to the maximum value specified by the manufacturer in order to protect the glass seals.

CHAPTER XVII

RECTIFIERS AND POWER SUPPLIES

1. Introduction.—Any device offering a greater impedance to the flow of electricity in one direction than it offers in the opposite direction will rectify an alternating current and give direct current. This may be accomplished by synchronous contactors, copper oxide cells, thermionic vacuum diodes, gas-discharge tubes, etc.

The thermionic vacuum diode is a very useful rectifier. On the half cycle of alternating voltage for which the thermionic element is negative the electrons emitted thermionically are attracted to the plate. However, on the following half cycle the cold plate, which cannot emit electrons, is negative, and no current can flow. If such a tube is placed in series with a source of alternating current and a load resistor, a pulsating unidirectional current will flow in the circuit. The voltage drop e_b , across a thermionic diode, during the conducting period depends on the current flowing according to the three-halves-power law. The waveform of current flowing during the conducting period for a resistance load may be determined graphically, Fig. 1.1, where the three-halves power relation is plotted with e_b as a function of i_b in the upper left. For any instant of time t_1 when the applied voltage is e_1 , indicated by the point P on the sine wave (lower left), the current is determined by the intersection Q of the resistance line for that voltage and the tube characteristic. The tube drop e_b and resistance drop i_1R are indicated. During the positive half cycle of applied voltage, the current increases slowly at first and then approaches the sinusoidal form. This slow increase causes the vacuum rectifier to produce less interference than the arc rectifier, which is sometimes a deciding factor in the selection of tubes.

Figure 1.2 shows the characteristic tube drops of several commonly used vacuum rectifiers. Tubes such as the 878 and 80 (5Y3-G) are considered high-impedance tubes, while the 83-v (5V4-G) is considered a low-impedance tube. The line marked Hg

indicates for comparison the approximately constant drop of the vapor rectifier tubes.

The vacuum rectifier can be built to operate safely with much higher inverse voltages than are possible with arc rectifiers. Since

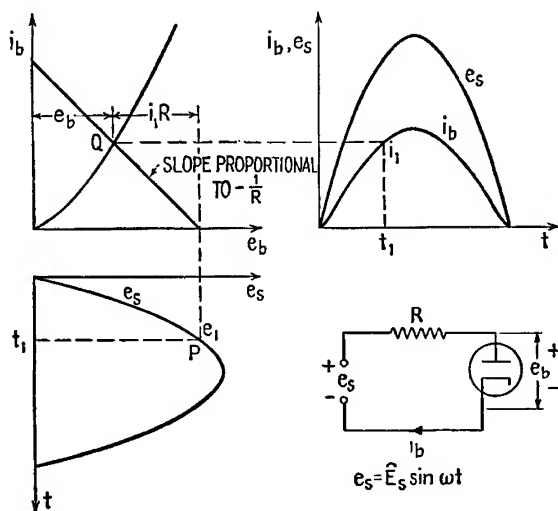


FIG. 1.1.—Determination of waveform for a vacuum thermionic rectifier.

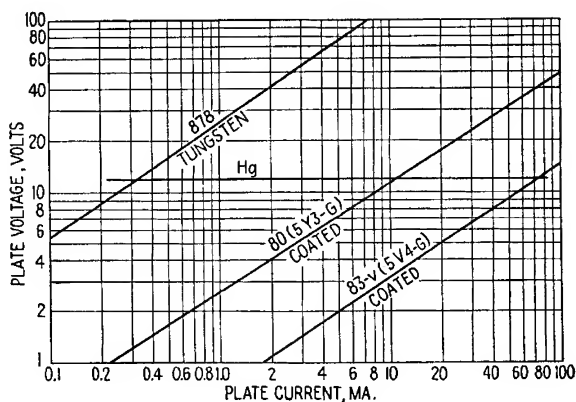


Fig. 1.2.—Characteristics of typical rectifier tubes.

the volt-ampere characteristic of the vacuum rectifiers is rising, a certain protection from overloads is inherent. When rectifiers having tungsten filaments are used, the finite emission of the filament provides a positive protection against extremely high short-circuit currents. The thermionic rectifier tubes are often constructed with

two anodes so that a single tube can be used in a full-wave rectifier circuit.

When arc tubes are used as rectifiers, a starting voltage somewhat higher than the normal arc drop must be reached by the applied voltage before current flows, causing an abrupt starting of the current, Fig. 1.3. This abrupt starting of current can cause objectionable interference, which can be radiated for considerable distances. In order to reduce this radiation the rectifier tube and circuit may be shielded and r-f chokes placed in the anode lead. When arc tubes are used to supply an oscillator, r-f chokes should

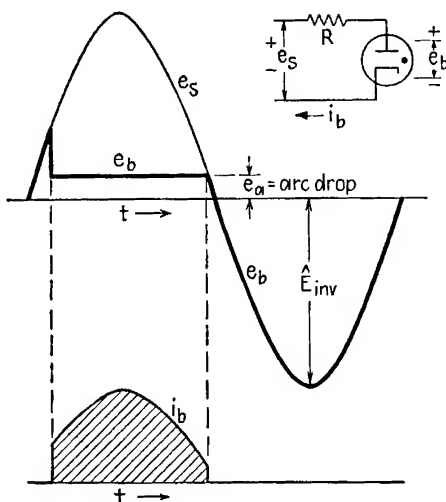


FIG. 1.3.—Waveform of current with an arc rectifier.

be placed in the d-c leads to protect the tube from r-f currents that might cause arc-backs.

2. Half-wave Rectifier Circuits.¹—In general, the tube drop of the rectifier is neglected in analyzing rectifier circuits, since it is usually small compared with the load voltage. The distortion in the current near zero, pointed out in Sec. 1, is also neglected, the current being assumed to start as soon as the applied voltage exceeds the load voltage.

Resistance Load.—In Fig. 2.1 a source of alternating voltage, such as the secondary of a transformer connected to the power line,

¹ J. D. COBINE, "Gaseous Conductors," Chap. XII, McGraw-Hill Book Company, Inc., 1941; J. MILLMAN and S. SEELY, "Electronics," Chap. XII, McGraw-Hill Book Company, Inc., 1941.

is assumed to have zero internal impedance, so that the voltage is independent of load. This voltage is

$$e_s = \sqrt{2} E_s \sin \omega t = \hat{E}_s \sin \omega t \quad (2.1)$$

where E_s is the effective (rms) value of the applied voltage and \hat{E}_s is its amplitude. During the half cycle when the anode is positive, the emf equation is

$$iR = e_s = \hat{E}_s \sin \omega t \quad 0 < \omega t < \pi \quad (2.2)$$

so that

$$i = \frac{\hat{E}_s}{R} \sin \omega t \quad (2.3)$$

During the negative half cycle, $i = 0$, Fig. 2.1. The average value

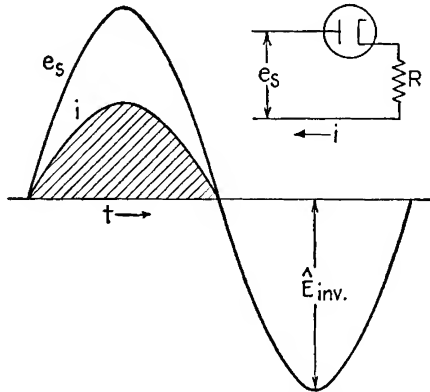


FIG. 2.1.—Rectifier relations for pure resistance load.

of current, indicated by a d-c ammeter, is

$$\bar{I} = \frac{1}{2\pi} \int_0^{2\pi} i d(\omega t) = \frac{1}{2\pi} \int_0^{\pi} \frac{\hat{E}_s}{R} \sin \omega t d(\omega t) \quad (2.4)$$

or

$$\bar{I} = \frac{1}{\pi} \frac{\hat{E}_s}{R} = 0.45 \frac{E_s}{R} \text{ average amperes} \quad (2.5)$$

The average value of the output, or load voltage is

$$\bar{E}_z = \bar{I}R = (\hat{E}_s/\pi) = 0.45E_s$$

The instantaneous load voltage in this circuit naturally has the same waveform as the load current. Thus the average rectified current in the resistance circuit is only 45 per cent of the rms cur-

rent E_s/R that would flow if the rectifier tube were short-circuited. The effective, or rms, value of the current, indicated by an a-c ammeter, is

$$I = \sqrt{\frac{1}{2\pi} \int_0^{2\pi} i^2 d(\omega t)} \quad (2.6)$$

$$= \sqrt{\frac{1}{2\pi} \int_0^\pi \left(\frac{\hat{E}_s}{R} \sin \omega t \right)^2 d\omega t} \quad (2.7)$$

or

$$I = \frac{1}{2} \frac{\hat{E}_s}{R} = 0.707 \frac{E_s}{R} \text{ rms amperes} \quad (2.8)$$

The peak value of the current, which must be safely passed by the rectifier tube, is

$$\hat{I} = 3.14 \bar{I} \quad (2.9)$$

Equations (2.5) and (2.9) may also be obtained from (10.9), Chap. IX. During the half cycle in which the anode is negative there is no voltage drop across the resistance so that the peak value of the applied voltage appears across the tube. The peak value of the voltage across the tube when it is not conducting is called the *peak inverse voltage*, \hat{E}_{inv} , Fig. 2.1. In this case $\hat{E}_{inv} = \hat{E}_s$. The rectifier tube must withstand \hat{E}_{inv} without a "cathode spot" being formed on the anode, which would remove the rectifying properties of the tube and cause what is known as an *arc-back*. Rectifier-tube data include the safe value of \hat{E}_{inv} for each tube as well as the average- and peak-current ratings.

Battery and Resistance Load.—When a single anode rectifier is used to charge a battery, with a resistance to control the current, Fig. 2.2, the emf equation during the positive half cycle of the applied voltage is

$$e_s = E_s \sin \omega t = iR + \bar{E}_B + e_a \quad (2.10)$$

where \bar{E}_B is the battery voltage. The arc drop e_a , or plate-to-cathode voltage during conduction, is included here since it is constant in magnitude for a gas tube and therefore like \bar{E}_B . From (2.10) the current is

$$i = \frac{\hat{E}_s}{R} \sin \omega t - \frac{\bar{E}_B + e_a}{R} \quad (2.11)$$

subject to the restriction that current can flow in only one direction so that i must be positive. The current can flow only during the interval between t_1 and t_2 , shown by the shaded area of Fig. 2.2,

for which e_s is greater than $\bar{E}_B + e_a$. In this case the peak inverse voltage is equal to the battery voltage plus the peak value of e_s , or

$$\hat{E}_{inv} = \bar{E}_B + \sqrt{2} E_s \quad (2.12)$$

Pure Capacitance Load.—When an a-c supply voltage e_s is applied to a rectifier and an uncharged capacitor in series at the instant the voltage is passing through zero to become positive,

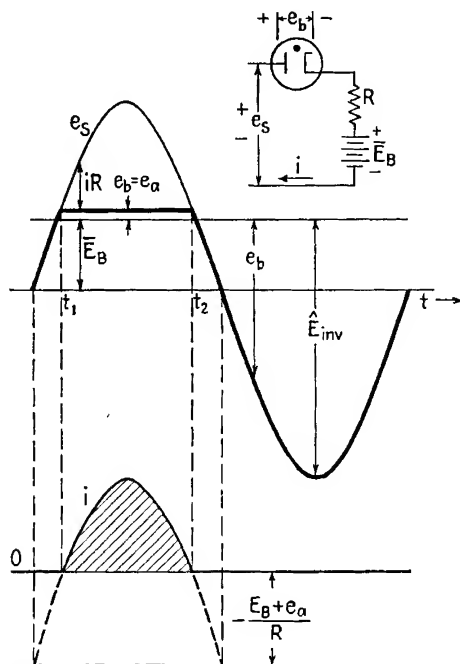


FIG. 2.2.—Voltage and current relations for a battery and resistance load.

Fig. 2.3, the current increases abruptly, neglecting tube drop, to the value it would have under steady-state conditions on alternating current with the rectifier short-circuited (dotted portion of sine wave). During the first quarter cycle the capacitor charges up to the peak value of the applied voltage \hat{E}_s , and then the tube ceases to conduct. When the applied voltage reverses, the tube has the sum of the capacitor voltage and the peak of the applied voltage. Thus the peak inverse voltage is $\hat{E}_{inv} = 2\hat{E}_s = 2\sqrt{2} E_s$.

The effect of switching an uncharged capacitor at an instant other than the normal zero of voltage may be seen by considering

the circuit of Fig. 2.4. In this figure R_s represents the source, or transformer secondary, resistance and is usually small. The dotted curve of Fig. 2.4 represents the normal current that would flow in the series RC circuit if the rectifier were short-circuited. When

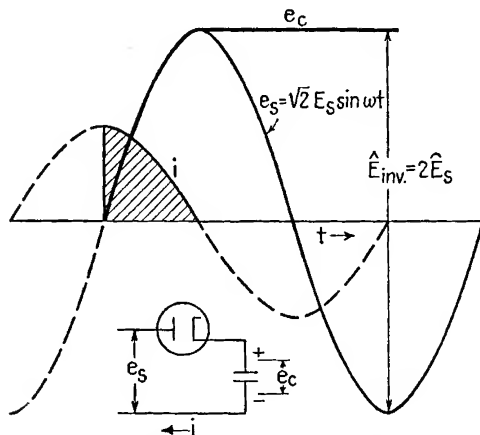


FIG. 2.3.—Rectifier with pure capacitance load.

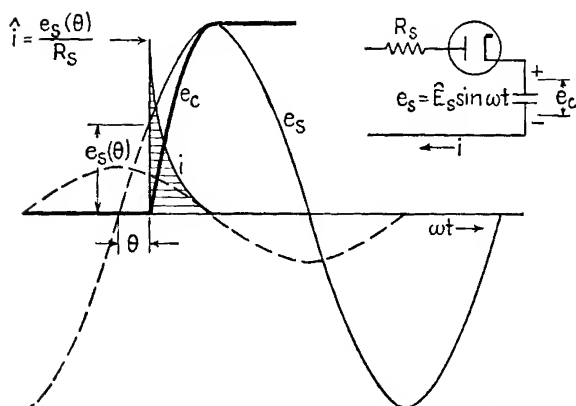


FIG. 2.4.—Effect of switching angle θ on transient current to an uncharged capacitor.

the voltage is switched on at the instant θ , it has a magnitude $e_s(\theta) = \hat{E}_s \sin \theta$. This voltage appears entirely across the resistance R_s at the first instant, since the capacitor is assumed initially uncharged. Thus the current surge reaches a peak value

$$\hat{i} = \frac{e_s(\theta)}{R_s}$$

Since R_s is usually small, this peak current can be excessive. Such a peak current can occur each cycle with a grid-controlled rectifier. The limited peak-current rating of the gas-discharge rectifiers *prohibits* their use with a capacitive load unless sufficient series resistance is added to limit the peak current to the peak-current rating of the tube.

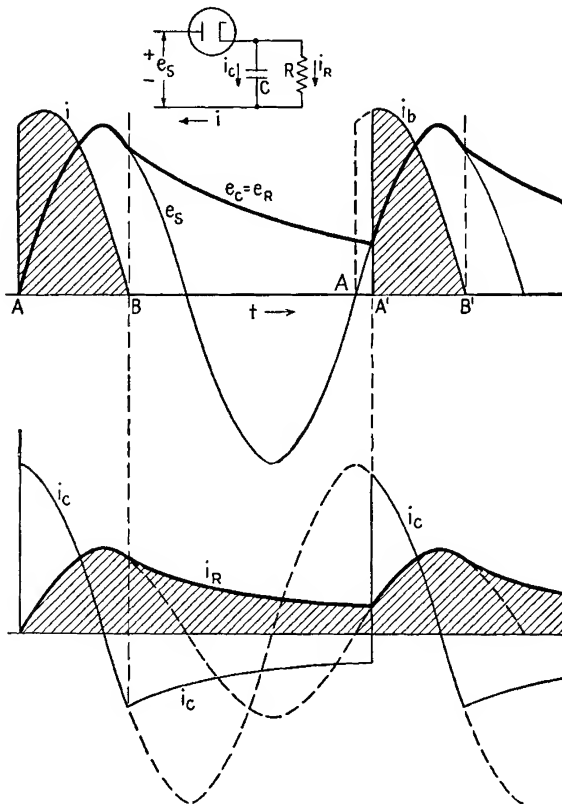


FIG. 2.5.—Current and voltage relations in load of R and C in parallel. Solid, heavy lines show rectifier-circuit quantities; dotted lines show normal steady-state quantities (rectifier short-circuited).

Parallel Load of R and C .—A capacitor charged by a rectifier acts as a simple filter and serves as a reservoir of charge to keep the voltage across a parallel load resistor relatively constant. The current and voltage waves for a load of R and C in parallel are shown in Fig. 2.5. The power is assumed to be switched on at A so that the first cycle represents the initial charging condition of the capaci-

tor. (If the switching had taken place later in the first positive half cycle, a switching transient would occur, as discussed in the preceding paragraph.) The currents flowing in R and C during the period the tube conducts have the same sinusoidal form as if the tube were short-circuited. The capacitance component i_c leads e_s by 90° , while the resistance component i_R is in phase with e_s . At the instant B , i_c is equal and opposite to i_R . At this instant the tube current $i_b = i_c + i_R$ is zero; and since the rectifier tube cannot carry a reverse current, conduction ceases. From B to A' the capacitor discharges in the usual exponential manner of an RC circuit. At A' the applied voltage becomes greater than the load

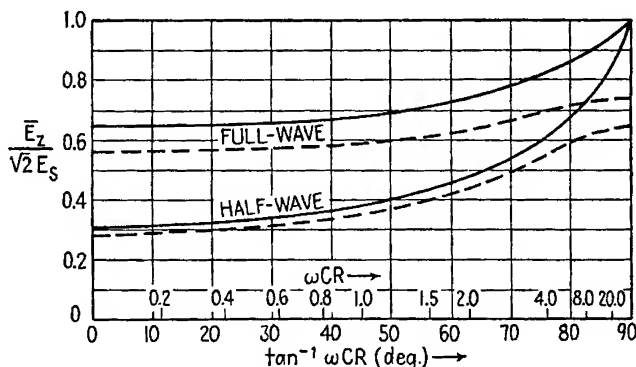


FIG. 2.6.—Relation between average load voltage and characteristic angle of parallel RC circuit supplied by a rectifier (average load voltage $\bar{E}_z = \bar{E}_R$ is referred to peak of applied voltage, $\sqrt{2} E_s$). Solid curves are for tube drop = 0; dotted curves are for load resistance/rectifier resistance = 10.

voltage, and the tube can conduct again. Because of the residual charge on C the instant A' is somewhat later in the cycle than A . All subsequent half cycles will be like the one starting at A' . An important disadvantage of this type of filter is the high peak current in the tube compared with the average value of the load current i_R . The greater the load current, *i.e.*, the lower the value of R , the greater the variation in the load voltage throughout the cycle.

Because of the importance of the RC load and the difficulty in calculating the various properties of the circuit, several figures due to Waidelich¹ are included. Figure 2.6 shows the relation between the average load voltage \bar{E}_z , referred to the peak a-c voltage, and the characteristic angle of the load, $\tan^{-1} \omega RC$. The solid curves of Fig. 2.6 neglect the tube drop of gas tubes and are reasonably good

¹ D. L. WAIDELICH, *Trans. A.I.E.E.*, **60**, 1161, 1941.

for vacuum rectifiers as long as the ratio of the load resistance to the tube resistance is greater than about 50. The dotted curves are for a ratio 10 of load resistance to tube resistance. The upper two curves refer to the characteristics of two rectifier tubes arranged so that both half cycles of the a-c supply contribute to the load, *i.e.*, a full-wave rectifier, Sec. 3. Figure 2.7 shows the effect of load angle on the peak current per tube and on the peak inverse voltage. Figure 2.8 shows the ratio of the effective, or rms, value of the ripple voltage, *i.e.*, the variational component of the load voltage, to the average value of the load voltage as function of the load angle. This ratio is often called the *ripple factor*. Figures 2.6 to

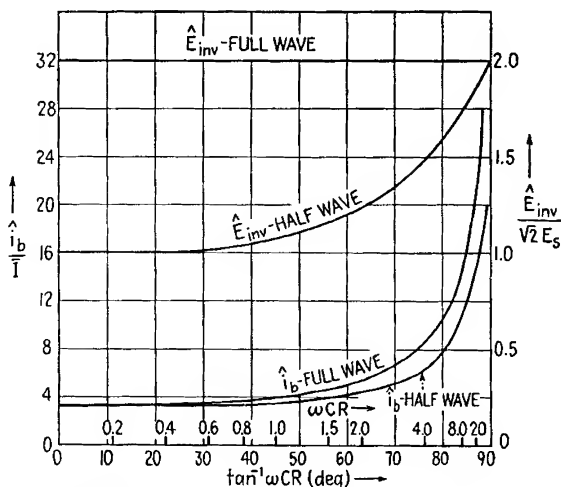


FIG. 2.7. — Peak voltages and currents for rectifier tube supplying parallel RC load.
 \bar{I} = average current per tube.

2.8 are useful in designing a load circuit to meet specific requirements. Load requirements will include the allowable ripple factor, load voltage, and current.

The steps in the design are as follows: The value of ωRC is found for the ripple factor specified, Fig. 2.8 being used; the value of R is given by the ratio of d-c load voltage to d-c load current; the rms transformer supply voltage E_s for each anode circuit is found from the ratio $\bar{E}/\sqrt{2} E_s$, Fig. 2.6, and the specified load voltage; and finally the peak tube current and peak inverse voltage are found from Fig. 2.7.

*Inductance Load.*¹—When a rectifier has a pure inductance load,

¹ COBINE, *op. cit.*, pp. 481–485.

Fig. 2.9a, the following voltage equation must be integrated to obtain the current wave:

$$e_s = \hat{E}_s \sin \omega t = \omega L \frac{di}{d(\omega t)} \quad (2.13)$$

The integration of (2.13) gives a current, (2.14), that is always positive as shown in Fig. 2.9a and that reaches a *peak* value that is

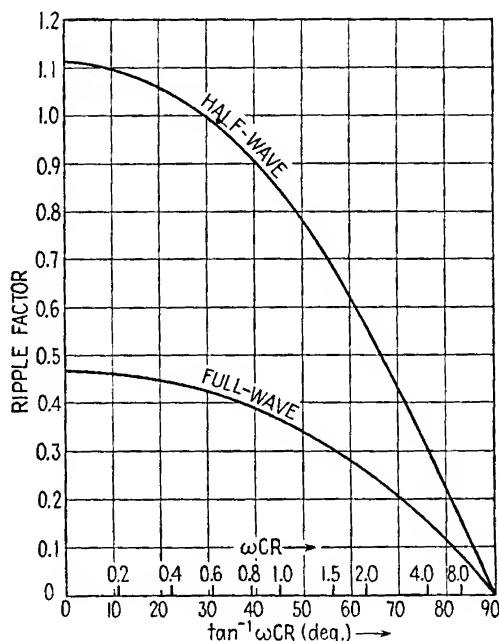


FIG. 2.8.—Ripple factor of parallel RC load as function of characteristic angle of load.

twice the peak current that would flow if the tube were short-circuited.

$$i = \frac{\hat{E}_s}{\omega L} (1 - \cos \omega t) \quad (2.14)$$

Conduction through the tube is continuous for 360° of the cycle. The *average* value of this current is obtained by dropping the $\cos \omega t$ term of (2.14), so that

$$\bar{I} = \frac{\hat{E}_s}{\omega L} = \frac{\sqrt{2} E_s}{\omega L} \quad (2.15)$$

which is the peak value of the steady-state current that would appear if the rectifier were short-circuited. Observe that the quantity $E_s/\omega L$ is the effective value of the current that flows when the

rectifier is short-circuited. It is important to note that a relay coil designed for use with an unrectified voltage of a given value may be saturated if the same a-c voltage is used with a grid-controlled rectifier to control the operation of the relay.

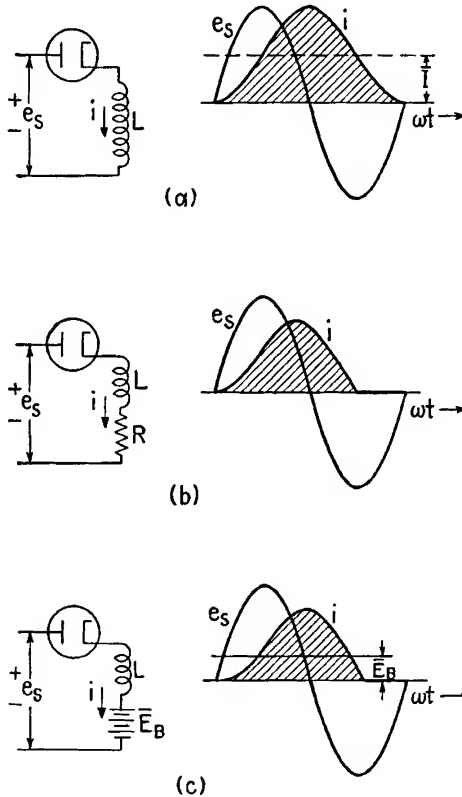


FIG. 2.9.—Half-wave rectifier with inductive load. (a) Pure inductance, $\sqrt{2} E_s / \omega L = 1$; (b) $\omega L / R = 4$, $\sqrt{2} E_s / \sqrt{R^2 + \omega^2 L^2} = 1$; (c) $E_B = \sqrt{2} E_s / 4$; $\sqrt{2} E_s / \omega L = 1$.

When a resistance is added to an inductance, the differential equation to be solved during the conducting period of the half-wave rectifier is

$$\omega L \frac{di}{d(\omega t)} + Ri = \hat{E}_s \sin \omega t \quad (2.16)$$

which may be integrated as

$$i = \frac{\hat{E}_s}{\sqrt{R^2 + \omega^2 L^2}} [\sin (\omega t - \phi) - \sin (-\phi) e^{-Rt/L}] \quad (2.17)$$

where

$$\phi = \tan^{-1} \frac{\omega L}{R}$$

This current is represented by the curve of Fig. 2.9b. The conduction period is seen to be less than 360° but more than 180° .

A battery may be charged more efficiently by using an inductance to limit the current instead of a series resistance. The differential equation for a half-wave rectifier and a series inductance and battery load is

$$\omega L \frac{di}{d(\omega t)} + \bar{E}_B = \bar{E}_S \sin \omega t \quad (2.18)$$

The solution of this equation is

$$i = \frac{\bar{E}_S}{\omega L} (1 - \cos \omega t) - \frac{\bar{E}_B}{\omega L} (\omega t) \quad (2.19)$$

which is the same as the addition of the current due to a rectifier with a pure inductance load and the linearly increasing current that results when a perfect battery is connected to a pure inductance. This current is shown in Fig. 2.9c.

3. Full-wave Rectifier.—In order to produce a smoother output wave than is possible with a half-wave rectifier, use is made of a

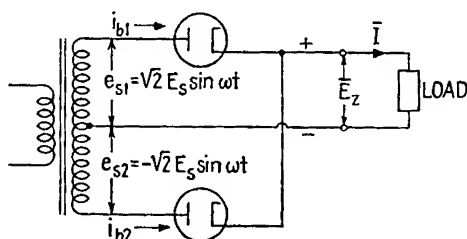


FIG. 3.1.—Full-wave rectifier circuit.

center-tapped transformer with the anodes of two tubes connected to the opposite ends of the transformer, Fig. 3.1. With this arrangement the anodes conduct on alternate half cycles and supply a pulsating voltage wave to the load, as shown by the heavy line of Fig. 3.2a. The average value \bar{E}_z of the load voltage is, from (10.10), Chap. IX,

$$\bar{E}_z = \frac{2}{\pi} \hat{E}_s = 0.9 E_s \quad (3.1)$$

double that of a half-wave rectifier. The average current to a pure resistance load is

$$\bar{I}_z = \frac{\bar{E}_z}{R} = 0.9 \frac{E_s}{R} \quad (3.2)$$

The waveforms of current obtained with R and with RL loads are shown in Fig. 3.2.

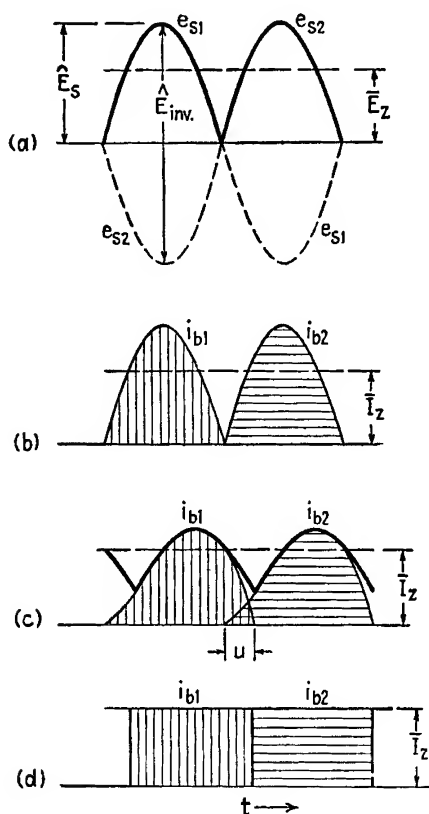


FIG. 3.2.—Full-wave-rectifier waveforms. (a) Voltages for pure resistance load; (b) current for pure resistance load; (c) current for load of R and L , u = period of overlap; (d) idealized current for highly inductive load with no inductance in anode leads.

When the load is a pure resistance, the load current wave is the same as the rectifier output voltage, Fig. 3.2b. The average plate current in each diode is $\bar{I}_b = \bar{I}_z/2$, and, from (10.9), Chap. IX, the peak current that must be passed by each diode for resistance load is

$$\hat{i}_b = \pi \bar{I}_b = \frac{\pi}{2} \bar{I}_z$$

where \bar{I}_z is the average value of the load current. Thus for each diode the ratio of peak plate current to average plate current is π , or 3.14.

When a very large inductance is used in series with the load resistance, the load current will remain constant at its average value. Each rectifier anode then supplies a square block of current shown by the shaded areas of Fig. 3.2*d*. In this case the peak

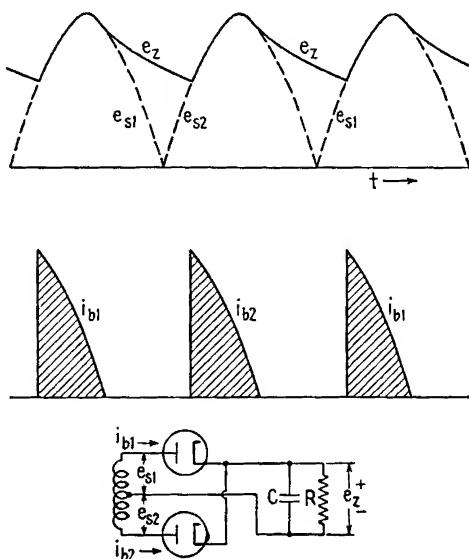


FIG. 3.3.—Waveforms of full-wave rectifier with parallel RC load.

current per diode is equal to the average load current; the average current per diode is as before, one-half the average load current. Most practical loads lie between the pure resistance load and the highly inductive load, and the current waveforms are similar to those of Fig. 3.2*c*.

If the tube is assumed to have no tube drop when conducting, the peak inverse voltage is

$$\hat{E}_{inv} = 2\hat{E}_s = 2\sqrt{2}E_s \quad (3.3)$$

or, from (3.1),

$$\hat{E}_{inv} = 2\hat{E}_s = \pi \bar{E}_z \quad (3.4)$$

The full-wave relations for an RC load, Fig. 3.3, may be derived

by using the same principles followed for that load with a half-wave supply.

Full-wave Bridge Rectifier.—The full-wave bridge circuit shown in Fig. 3.4 is often used for high voltages because of its lower peak inverse voltage for a given load voltage. The inverse voltage is the load voltage. One filament transformer may be used for the two filaments connected together, but the other two cathodes, being at different potentials, must be heated by separate transformers. During the half cycle when the point *a* is positive and *c* negative, the current flows through the circuit in the order *a-d-e-b-c*. On the following half cycle the path is *c-d-e-b-a*. Thus the point *e* is always positive with respect to point *b*, and the rectification is full wave. In this circuit the peak inverse voltage per tube is

$$\hat{E}_{inv} = \hat{E}_s = \frac{\pi}{2} \bar{E}_z \quad (3.5)$$

and the average voltage is

$$\bar{E}_z = 0.9E_s \quad (3.6)$$

as in (3.1). It should be noted that this is not really a bridge circuit so far as current flow is concerned but merely has the superficial appearance of a bridge. This circuit makes more efficient use of the

transformer than does the ordinary full-wave circuit of Fig. 3.1, since the entire secondary winding of Fig. 3.4 is used each half cycle, carrying an alternating current.

4. Voltage-doubling Rectifier Circuits.—In the circuit of Fig. 4.1 the capacitors are alternately charged by their respective diodes, while the load voltage is equal to the sum of the voltages of the two capacitors. This circuit is often used in small transformerless radios and in power-pulse generating circuits. The waveforms¹ of this circuit are similar to those of a full-wave rectifier with a load of *R* and *C* in parallel and are shown in Fig. 4.2. In this figure *i_{b1}* and *i_{b2}* are the plate currents of tubes *T₁* and *T₂*, charging capacitors *C₁* and *C₂*, respectively. Because of the high peak tube current, poor use is made of the current capacity of the tubes. The circuit has poor voltage regulation. For this reason it is generally used

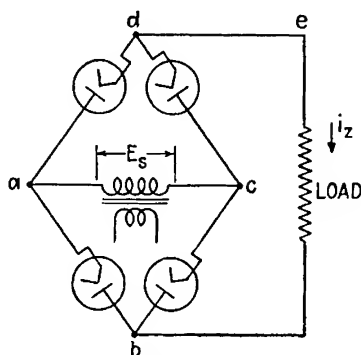


FIG. 3.4.—Full-wave bridge rectifier circuit.

¹ D. L. WAIDELICH, *Proc. I.R.E.*, **29**, 554, 1941.

where the load is constant. As the load current is reduced, the average output voltage approaches twice the peak value of the applied voltage, Fig. 4.3,¹ and the peak inverse voltage per tube approaches $2\hat{E}_s = 2\sqrt{2}E_s$. Electrolytic capacitors should not be used in this circuit unless $\omega CR > 4.79$, where $C = C_1 = C_2$,

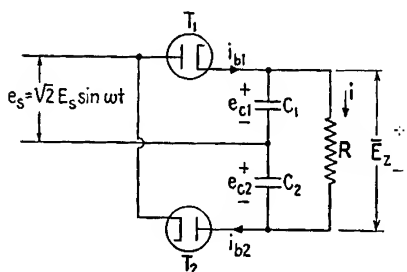


FIG. 4.1.—Full-wave-voltage-doubler circuit.

since for smaller values of ωCR the voltage of each capacitor reverses for part of the cycle. Figure 4.4¹ shows the relation between peak current, peak inverse voltage, ripple factor, and ωCR .

Figure 4.5² shows a half-wave voltage doubler in which the negative side of the load is returned to one of the a-c conductors.

This arrangement makes possible a common ground between load and supply. When T_2 conducts, capacitor C_1 charges to approximately the peak of the applied a-c voltage. Tube T_1 starts conducting when T_2 stops, and C_1 is discharged into the parallel load of C_2 and R . Figure 4.6² shows output voltage and ripple factor

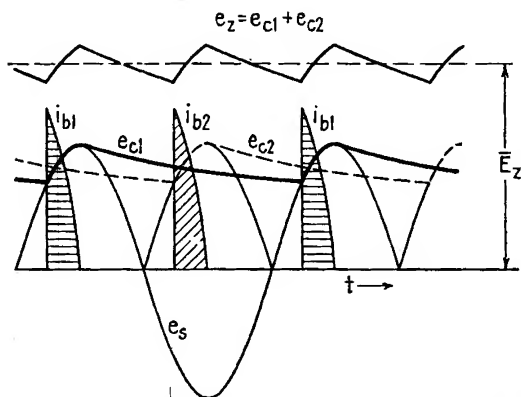


FIG. 4.2.—Voltage and current waveforms for full-wave voltage doubler.

as functions of ωCR , where $C = C_1 = C_2$. Peak-current and peak-inverse-voltage relations are shown in Fig. 4.7.² Figures 4.6 and 4.7 can be used to design a power supply to fit particular con-

¹ D. L. WAIDELICH, *Proc. I.R.E.*, **29**, 554, 1941.

² D. L. WAIDELICH and C. H. GLEASON, *Proc. I.R.E.*, **30**, 535, 1942.

ditions. The input capacitor C_1 must withstand the peak of the applied a-c voltage, while the load capacitor C_2 must withstand a maximum of twice the peak of the applied voltage. A polarized

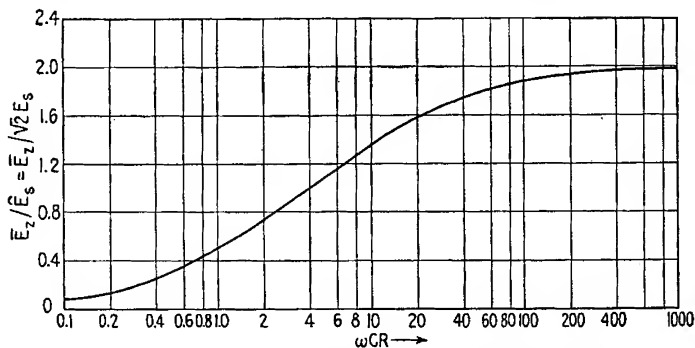


FIG. 4.3.—Ratio of average output voltage of full-wave voltage doubler to peak a-c, as function of load ωCR ; ω = supply fundamental, $2\pi f$.

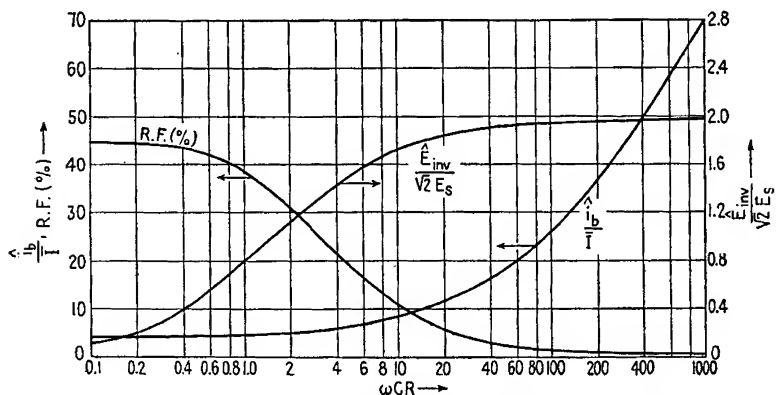


FIG. 4.4.—Full-wave-voltage-doubler characteristics. R.F. = ripple factor; \hat{E}_{inv} = peak inverse voltage; \hat{i}_b = peak current per tube; \bar{I} = average load current; E_s = rms applied voltage.

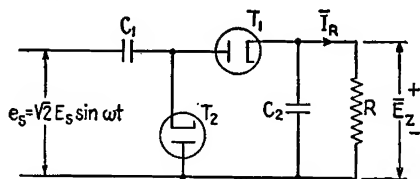


FIG. 4.5.—Half-wave-voltage-doubling rectifier.

electrolytic capacitor cannot be used for C_1 unless ωCR is greater than 4.5, because for lower values the voltage of this capacitor reverses for a part of the cycle. Capacitor C_2 may be electrolytic.

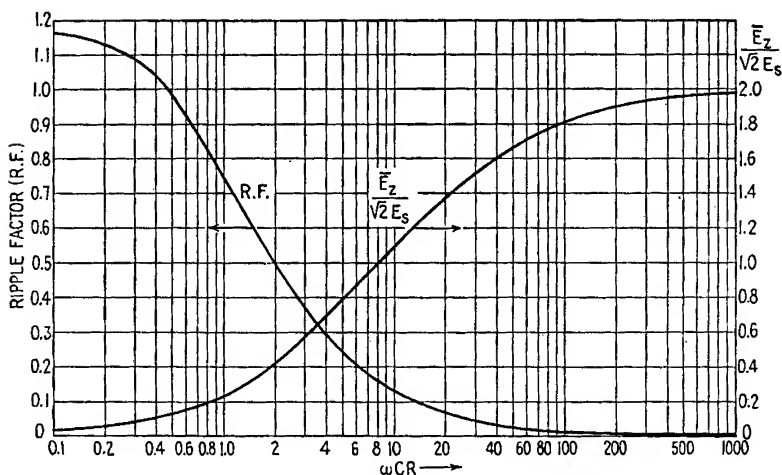


FIG. 4.6.—Average voltage and ripple factor of half-wave voltage doubler; $C = C_1 = C_2$.

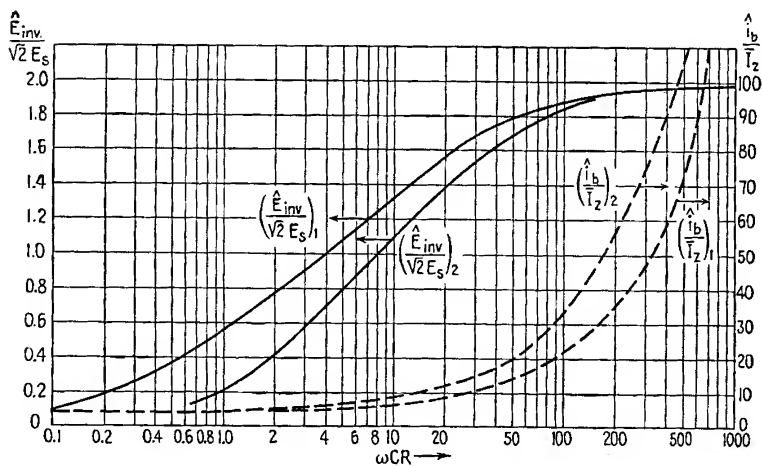


FIG. 4.7.—Peak inverse voltage and peak tube currents of half-wave voltage doubler. E_s = rms a-c supply voltage; \hat{E}_{inv} = peak inverse voltage per tube; \hat{i}_b = peak tube current; \bar{I}_2 = average load current; subscripts 1 and 2 refer to tubes T_1 and T_2 .

The normal operating range of both voltage-doubling circuits is for ωCR greater than 10. In the operating range the full-wave doubler, Fig. 4.1, has lower peak tube currents, slightly less ripple (at twice the frequency), and a slightly better voltage regulation than the half-wave circuit, Fig. 4.5. The peak inverse voltage

is somewhat lower for the half-wave circuit than for the full-wave circuit.

5. Polyphase Rectifiers.¹—When it is necessary that a rectifier furnish power in excess of 1 kilowatt, as for a high-power transmitter, a three-phase a-c supply and a polyphase arrangement of tubes are generally used. Polyphase rectifier circuits provide a very much smoother output-voltage wave than single-phase half-wave and full-wave circuits do. This may be seen by considering the harmonic components of the waves given in Table 6.1, Sec. 6.

The principles of polyphase rectifiers may be understood by considering the three-phase half-wave circuit shown in Fig. 5.1. The primary windings are connected in delta, and the secondary windings are in wye. The secondary winding 1 is on the same part of the transformer magnetic circuit as primary winding I. Sim-

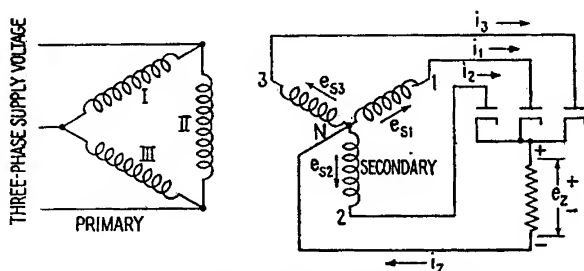


FIG. 5.1.—Three-phase half-wave rectifier.

ilarly, secondary winding 2 is associated with primary II and secondary winding 3 with primary III. The three groups of primary and secondary windings may be three single-phase transformers, or they may be placed on the separate legs of a three-phase transformer.

For satisfactory operation the primary windings must be connected in delta, since only one secondary winding carries current at a time. Consider the period during which anode 1 conducts, with zero current in secondaries 2 and 3. The secondary ampere turns in 1 must be balanced by an equal number of ampere turns in I. However, if the primary is in wye the current in I must flow in either II or III or in both in order to reach the primary supply lines. Since anodes 2 and 3 are idle, there are no balancing ampere

¹ MILLMAN and SEELY, *op. cit.*, pp. 381–412; COBINE, *op. cit.*, pp. 488–504; H. RISSIK, "The Fundamental Theory of Arc Convertors," Chapman & Hall, Ltd., 1939.

turns in the secondary for currents in II and III. Therefore the wye connection should not be used on the primary side. When a delta-connected primary is used, the primary current in I, because of the conduction of anode 1, can flow directly from one power line to another through the single winding I.

The secondary windings supply voltages e_{s1} , e_{s2} , e_{s3} , measured from the neutral N . These voltages which are applied to the anodes of the rectifier tubes are 120 electrical degrees apart and are defined by the following equations:

$$\begin{aligned}e_{s1} &= \sqrt{2} E_s \sin \omega t \\e_{s2} &= \sqrt{2} E_s \sin (\omega t - 120^\circ) \\e_{s3} &= \sqrt{2} E_s \sin (\omega t - 240^\circ)\end{aligned}$$

where E_s is the rms value of the voltage of each secondary winding. Figure 5.2*a* shows graphically the relation among the three anode voltages. The heavy curve e_z represents the output voltage of the rectifier. During the interval between A and B , the voltage of anode 1, e_{s1} , is the highest of the phase voltages so that this anode carries the entire load. In this period the other anodes are negative relative to the load and cannot conduct. At the instant B the voltage of anode 2, e_{s2} , becomes equal to that of anode 1. Since e_{s1} is decreasing at B while e_{s2} is increasing, the load is taken over by anode 2. At C , anode 3 assumes the load until the instant A' , when the cycle is repeated. Thus the load-voltage wave is the envelope of the three phase voltages. The fundamental ripple frequency of this circuit is three times the power frequency, and the lowest harmonic present is the third harmonic of the power frequency. Table 6.1 gives the rms values of the harmonics present in the output wave. If the load is a pure resistance, the load current is as shown in Fig. 5.2*b*. The portion of the current carried by each anode is indicated by the shaded areas. It is evident from this figure that for a resistance load each anode must carry a peak current that is greater than the average load current and that the average current carried by each anode is one-third the average load current. If the load is highly inductive, the load current will be held nearly constant at its average value, so that the anode-current waves approach as a limit the square waves shown in Fig. 5.2*c*. Table 7.1 gives the numerical characteristics of this rectifier circuit. In an actual rectifier the load voltage will be decreased slightly from the calculated values by the tube drop and by the effect of the transformer

reactance. The transformer leakage reactance prevents sudden changes in the anode current.¹

The delta-wye connection has the disadvantage of causing magnetic saturation of the transformer core. This is due to the unidirectional nature of the currents flowing in the secondary

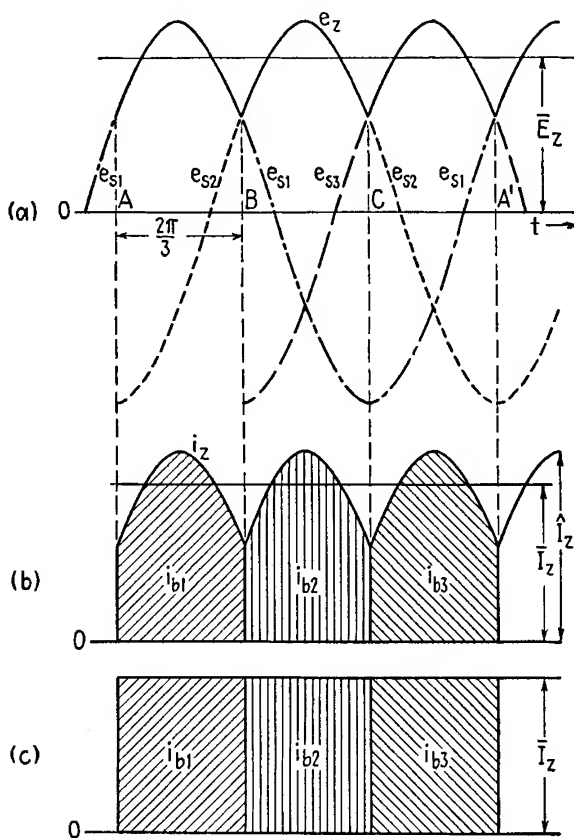


FIG. 5.2.—Characteristics of three-phase half-wave rectifier. (a) Transformer secondary and load voltages; (b) anode and load currents with resistance load; (c) anode and load currents with highly inductive load.

windings. This saturation can be avoided if each of the secondary windings consists of two sections. By reversing the connections of one section of each secondary phase, a delta-star rectifier connection using 6 anodes is obtained, Fig. 5.3. Secondary winding sections 1 and 4 have I as their primary winding. Similarly, the primary for

¹ COBINE, *op. cit.*, pp. 494–498.

sections 3 and 6 is II, and that for sections 2 and 5 is III. The 6 anode voltages produced by this 6-phase connection are 60 electrical degrees apart. Each anode conducts for 60 electrical degrees and must pass a peak current equal to the peak load current and an average current equal to $\bar{I}_z/6$. This connection results in a ripple frequency 6 times the fundamental supply frequency. Thus the lowest harmonic present is the 6th harmonic of the power supply, and the next lowest harmonic is the 12th. These harmonics have small amplitudes so that the filtering problem is greatly simplified.

Usually a rectifier of large output will use one of several possible connections¹ giving the same output wave as the six-phase rectifier but so arranged as to utilize the transformer capacity to greater

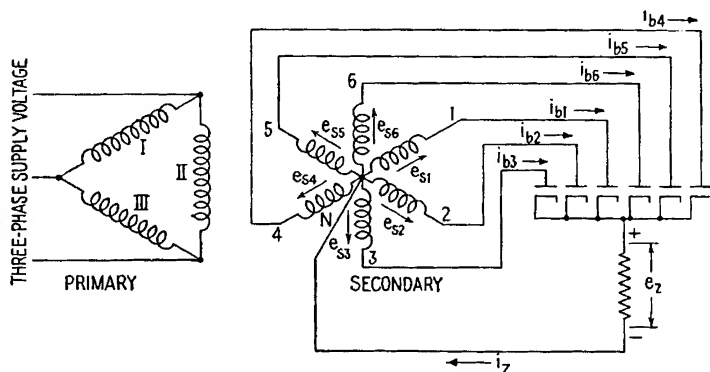


FIG. 5.3.—Six-phase half-wave rectifier.

advantage. If the rectifier is to produce high d-c voltages at relatively low currents, vacuum thermionic-rectifier tubes are used. However, if considerable current is desired at moderately high voltages, the steel-tank mercury-pool rectifier is widely used. This rectifier² consists of a steel tank containing a mercury pool that serves as the cathode for one or more anodes. The steel-tank rectifier can be made for very large output. The mercury-pool cathode can supply very heavy currents for brief periods without damage.

6. Harmonic Content of Rectified Voltages.³—The output voltage of a rectifier circuit can be represented by the method of Fourier analysis as being made up of a constant voltage equal

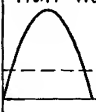
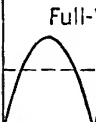
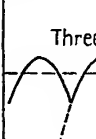
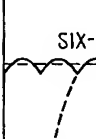
¹ *Ibid.*, pp. 500–505; MILLMAN and SEELY, *op. cit.*, pp. 384–405.

² COBINE, *op. cit.*, pp. 437–445.

³ L. B. W. JOLLEY, "Alternating Current Rectification," p. 18, John Wiley & Sons, Inc., 1928.

TABLE 6.1.—EFFECTIVE VALUES OF HARMONICS OF RECTIFIED WAVES

 Numerical values are the ratio $k_n = \frac{E_n}{E_s}$.

HARMONIC (n)	1*	2	3	4	6	8	9	10	12
FUNCTION	$+\sin \omega t$	$-\cos 2\omega t$	$-\cos 3\omega t$	$-\cos 4\omega t$	$-\cos 6\omega t$	$-\cos 8\omega t$	$-\cos 9\omega t$	$-\cos 10\omega t$	$-\cos 12\omega t$
WAVE FORM									
Half-Wave 	1.11	0.471	---	0.0943	0.0404	0.0224	---	0.0143	0.0099
Full-Wave 	---	0.471	---	0.0944	0.0405	0.0225	---	0.0143	0.0099
Three-Phase 	---	---	0.177	---	0.0405	---	0.0177	---	0.0099
SIX-PHASE 	---	---	---	---	0.0405	---	---	---	0.0099

* Fundamental Power Frequency

 E_s = rms a-c supply voltage per anode

to the average value of the wave plus an infinite series of harmonic voltages of decreasing magnitudes. Thus the analysis of the output voltage of a half-wave rectifier is, (10.9) Chap. IX,

$$e = \frac{1}{\pi} \hat{E}_s \left(1 + \frac{\pi}{2} \sin \omega t - \frac{2}{1 \cdot 3} \cos 2\omega t - \frac{2}{3 \cdot 5} \cos 4\omega t - \frac{2}{5 \cdot 7} \cos 6\omega t \cdots \right) \quad (6.1)$$

and the output voltage of a full-wave rectifier is, (10.10) Chap. X,

$$e = \frac{2}{\pi} \hat{E}_s \left(1 - \frac{2}{1 \cdot 3} \cos 2\omega t - \frac{2}{3 \cdot 5} \cos 4\omega t - \frac{2}{5 \cdot 7} \cos 6\omega t \cdots \right) \quad (6.2)$$

where ω is the angular frequency of the alternating-current source being rectified. The harmonic voltages of these and other rectified waves are listed in Table 6.1. The current that flows in a circuit when a rectified voltage is applied can be determined by calculating by the simple methods of steady-state a-c analysis the currents that flow due to each of the harmonics separately and adding these components to obtain the resultant current wave. Thus if E_n is the effective, or rms, value of the n th harmonic of a rectified voltage wave having a fundamental power frequency $f = \omega/2\pi$ obtained from Table 6.1, the rms current I_n due to that harmonic alone is

$$|I_n| = \frac{|E_n|}{|Z_n|}$$

where $|Z_n|$ is the magnitude of the impedance of the circuit to the frequency $n\omega$. In most rectifier analysis it is usually sufficient to calculate the rms current flowing due to one or two of the lowest

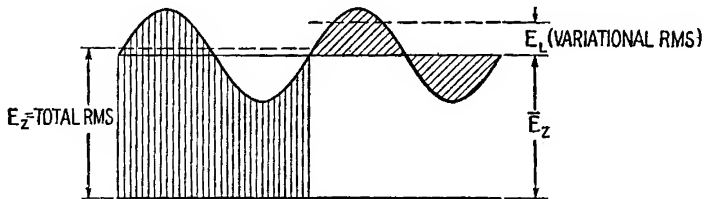


FIG. 6.1.—Generalized voltage output of rectifier, and smoothing circuit.

order harmonics, since the magnitude of the higher harmonics decreases rapidly. The pulsating output voltage of rectifiers is usually applied to the load through a network designed to absorb as much as possible of the a-c component. The voltage applied to the load may be generalized as shown in Fig. 6.1. In this figure, \bar{E}_z is the average value of the voltage, indicated by a d-c voltmeter, and E_z is the effective value of the output voltage (vertical shading), indicated by an rms a-c voltmeter. If a large capacitor is placed in series with an a-c voltmeter to block off the d-c component, only the variational component E_z will be measured (crosshatched cycle).

The term *ripple factor* (R.F.) is useful in characterizing a pulsating wave; it is defined as

$$\text{R.F.} = \frac{\text{rms value of all a-c components}}{\text{average value}} \quad (6.3)$$

Values of the ripple factor for three rectifier circuits are given in

Table 7.1. It is often sufficient as a first approximation to take the ripple factor equal to the relative value k_n (Table 6.1) of the lowest harmonic present. The ripple factor must be reduced to about 0.0005 for the microphone circuit of a transmitter, to about 0.0001 to 0.001 for an audio amplifier, and to less than 0.01 for a cathode-ray oscillograph beam supply.

7. Transformer Utilization Factor.¹—Transformers used in rectifier service are not loaded with sinusoidal voltages and currents as is assumed in most transformer-design theory. Rectifiers have nonsinusoidal currents whose I^2R heating effects in the transformer winding are greater than for sinusoidal currents of the same average value. The rating of the transformer is determined by the heat that the surface of the transformer can dissipate. For this reason, transformers are rated in volt-amperes rather than in watts of load. Since the currents flowing in the primary and in the secondary windings of a rectifier transformer usually have different wave-forms, a different factor must be used in determining the ratings of each of these windings. It is usually possible to use an average of the primary and secondary rating factors in determining a suitable size of transformer for a specific application. The *utilization factor*, or *utility factor* (U.F.), of a transformer winding is the ratio of the rectified output power to the total volt-ampere rating of the winding,

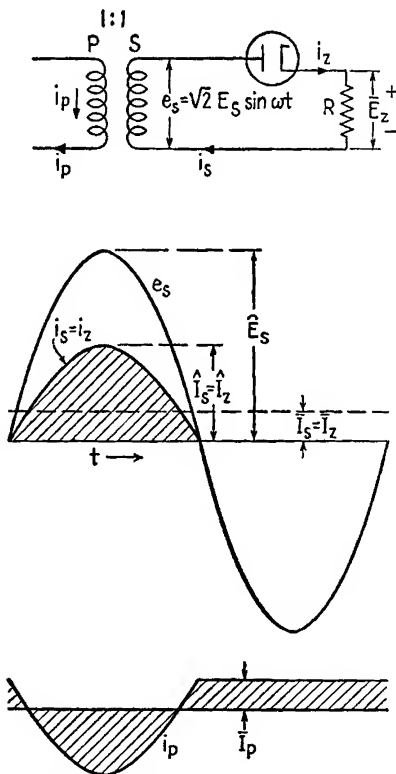


FIG. 7.1.—Primary and secondary currents of a rectifier transformer.

¹ F. J. TEAGE and J. F. GILL, "Mercury Arcs," p. 71, Methuen & Co., Ltd., H. RISSIK, "Mercury Arc Current Convertors," p. 25, Sir Isaac Pitman & Sons, Ltd., 1935.

$$\text{U.F.} = \frac{\text{total d-c power}}{\text{total volt-ampere rating of winding}} = \frac{\bar{E}_z \bar{I}_z}{pEI} \quad (7.1)$$

where p is the number of winding sections¹ ($p = 2$ for full-wave single-phase secondary), E is the rms voltage of the winding section, and I is the rms current in the winding.

The concept of utilization factor may be understood better by deriving it for the simple case of a half-wave rectifier with a resistance load supplied by a one-to-one ratio transformer. The transformer voltage and current waves are shown in Fig. 7.1. The average value over the entire cycle of the load current (which flows only during the interval 0 to π) is

$$\bar{I}_z = \frac{1}{\pi} \hat{I}_z$$

where \hat{I}_z is the peak value reached by the load current. The average load voltage is

$$\bar{E}_z = \frac{1}{\pi} \hat{E}_s$$

where \hat{E}_s is the amplitude of the a-c supply voltage. The d-c power of the load is

$$P = \bar{E}_z \bar{I}_z = \frac{1}{\pi^2} \hat{E}_s \hat{I}_z = 0.101 \hat{E}_s \hat{I}_z \quad (7.2)$$

The copper loss of the secondary winding is determined by the square of the rms value I of the current over a cycle, as given by

$$I_z^2 = \frac{\hat{I}_z^2}{2\pi} \int_0^\pi \sin^2 \theta d\theta = \frac{\hat{I}_z^2}{4} \quad (7.3)$$

In normal a-c use a sinusoidal current flows during the complete cycle. The rms value of such a current is the peak value divided by $\sqrt{2}$. If \hat{I}_z is the peak value of a sinusoidal alternating current that will cause the same loss and the same temperature rise in the secondary winding as the pulsating current of the rectifier,

$$\frac{\hat{I}_z^2}{4} = \frac{\hat{I}_z^2}{2}$$

¹ For the secondary, p is usually equal to the number of anodes. The reader is referred to H. Rissik, "The Fundamental Theory of Arc Convertors," Chapman & Hall, Ltd., 1939, for special cases in polyphase rectification.

Then the equivalent sinusoidal current has an amplitude

$$\hat{I}_x = \frac{\hat{I}_z}{\sqrt{2}} \quad (7.4)$$

The volt-ampere rating of the transformer winding operating at its normal temperature is given by the product of the rms voltage and the rms value of the equivalent current of (7.4).

$$VA = \frac{\hat{E}_s}{\sqrt{2}} \frac{\hat{I}_x}{\sqrt{2}} = \frac{\hat{E}_s \hat{I}_z}{2 \sqrt{2}} = 0.354 \hat{E}_s \hat{I}_z \quad (7.5)$$

By the definition of utilization factor, (7.1), and by (7.5),

$$\text{U.F.} = \frac{P}{VA} = \frac{0.101 \hat{E}_s \hat{I}_z}{0.354 \hat{E}_s \hat{I}_z} = 0.285 \quad (7.6)$$

Thus the secondary winding of the transformer supplying the half-wave rectifier must have a rating of $1/\text{U.F.} = 3.49$ times that required for the same power with a sinusoidal current. Since no direct current passes between transformer secondary and primary windings, the primary current wave is as shown by i_p in Fig. 7.1, where the zero line corresponds to the average value of the secondary current. This gives a value of 0.371 for the U.F. of the primary winding. The unidirectional current in the secondary produces a d-c magnetization of the transformer core, which can cause an excessive increase in the exciting current. This means that a specially designed transformer having, in effect, a U.F. considerably smaller than 0.371 should be used. If a transformer designed for sine-wave operation is used, it may be necessary to use one designed for a higher primary voltage in order to reduce the exciting current for the actual voltage to a reasonable value.

A full-wave rectifier having a highly inductive load, such as is provided by a choke-input filter, will subject the secondary of the transformer to pulsating square waves of current. This will result in an alternating square wave of current in the primary. Under this condition of operation the U.F. of the secondary is 0.637, while that of the primary is 0.9.

Commercial a-c transformers for ordinary power and lighting are designed to have approximately equal copper losses (I^2R) in primary and secondary windings with sinusoidal voltages and currents. When a transformer is built for rectifier service, the secondary copper is increased over the normal amount. If a transformer originally designed for normal a-c operation is to be used in rectifier service, an average utilization factor equal to the average of the

primary and secondary utilization factors for the rectifier circuit in question may be used. This will determine a rating that will somewhat overload the secondary winding and underload the primary winding but will not overheat the entire transformer. It should be noted that in the use of transformers already built, these values of utilization factor serve as a guide in selecting the proper current rating. The actual heating of a transformer so selected will depend on how conservative a design is followed.

TABLE 7.1.—RECTIFIER CONSTANTS

Rectifier circuit	Half-wave	Full-wave	Full-wave bridge	Three-phase	Six-phase
Average voltage $= \frac{\bar{E}_z}{E_s}$	0.45	0.9	0.9	1.17	1.35
Rms anode supply voltage					
Utilization factor:					
Primary	0.371†	0.9‡	0.9‡	0.827‡	0.78‡
Secondary	0.285†	0.637‡	0.9‡	0.675‡	0.552‡
Peak inverse voltage* $= \frac{\hat{E}_{inv}}{E_s}$	3.14	3.14	1.57	2.09	2.09
Average voltage $= \frac{\bar{E}_z}{E_s}$					
Peak inverse voltage* $= \frac{\hat{E}_{inv}}{E_s}$	1.41	2.82	1.41	2.44	2.48
Rms anode supply voltage					
Average anode current	1.0	0.5	0.5	0.33	0.167
Average load current					
Peak anode current	2.0‡	1.0‡	1.0‡	1.0‡	1.0‡
Average load current	3.14†	1.57†	1.57†	1.21†	1.05†
Ripple factor	1.22	0.484	0.484	0.182	0.042

* Greatest possible value.

† Resistance load.

‡ Large choke assumed.

8. Power Filters.¹—Simple filters such as a capacitor paralleling the load and an inductance in series with the load have already been discussed. When a smooth output-voltage wave is desired over a considerable range of current, it is necessary to make use of somewhat more elaborate methods of smoothing. In general, the filter should absorb a maximum of the developed variational voltage and provide a by-pass around the load for such alternating current as may flow.

¹ MILLMAN and SEELY, *op. cit.*, Chap. XIV; F. E. TERMAN, "Radio Engineering," 2d ed., Chap. XI, McGraw-Hill Book Company, Inc., 1937; H. J. REICH, "Theory and Application of Electron Tubes," 2d ed., Chap. XIV, pp. 582-590, McGraw-Hill Book Company, Inc., 1944.

Choke-input Filters.—In electronic applications, one- and two-section choke-input filters, Fig. 8.1, are used for polyphase rectifier circuits, for *all* gas-tube rectifiers, and in general when high power is required with a high ratio of average to peak value of tube current. A very useful approximate analysis of this type of circuit is based on the fact that for a good filter ωL is very much greater than $1/\omega C$. The output voltage of the rectifier can be expressed as

$$e = \bar{E} + \sum_p^n \sqrt{2} E_n \cos n\omega t \quad (8.1)$$

where E_n is the effective value of the n th harmonic (Table 6.1) and p is the number of anodes of the rectifier firing separately ($p = 1$ for half wave, $p = 2$ for full wave). The currents and voltages at any point can be determined for each harmonic component separately by the usual method of steady-state a-c analysis. If the resultant wave is desired, these harmonics can then be added in their correct

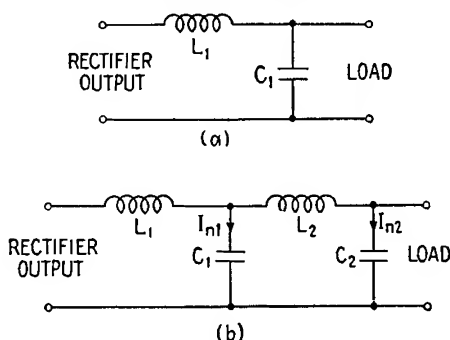


FIG. 8.1.—Choke-input filters: (a) one-section and (b) two-section.

time relation. Frequently calculations for a few of the lowest harmonics are adequate.

In analyzing the two-section choke-input filter it will be assumed at first that practically all the alternating voltage due to the n th harmonic appears across the first choke L_1 and all the alternating current flows through the capacitor C_1 . The rms current in C_1 is

$$I_{n1} \doteq \frac{E_n}{n\omega L_1 - \frac{1}{n\omega C_1}}$$

or since L_1 and C_1 are usually large in a power filter,

$$I_{n1} \doteq \frac{E_n}{n\omega L_1} \quad (8.2)$$

where E_n is the effective value of the n th harmonic of input voltage to the filter. This harmonic current produces a voltage drop¹ in C_1 of

$$E_{n1} = \frac{I_{n1}}{n\omega C_1} = \frac{E_n}{n^2\omega^2 L_1 C_1} \quad (\text{rms}) \quad (8.3)$$

This is the output ripple voltage of the n th harmonic for a single-section filter. The same procedure is followed for the second section. The a-c voltage across the second capacitor is assumed to be negligible compared with the a-c voltage across the second choke, so that the current in C_2 due to the n th harmonic of the input voltage is

$$I_{n2} \doteq \frac{E_{n1}}{n\omega L_2} \quad (8.4)$$

where E_{n1} is the harmonic voltage impressed on the second section by the first-section capacitor, (8.3). Then

$$E_{n2} = \frac{E_{n1}}{n^3\omega^3 L_1 C_1 L_2} \quad (8.5)$$

and the voltage developed across C_2 due to this harmonic current is

$$E_{n2} = \frac{I_{n2}}{n\omega C_2} = \frac{E_n}{n^4\omega^4 L_1 L_2 C_1 C_2} \quad (\text{rms}) \quad (8.6)$$

and is the harmonic voltage developed across the load. Since the rms value of the rectifier output voltage of the n th harmonic drops off very rapidly as n increases, it is often sufficient to consider the ripple factor as given by the ratio of the lowest harmonic voltage across C_2 to the average value. Since the lowest harmonic present is equal to the number of anodes p firing separately in the rectifier ($p = 2$ for a single-phase full-wave rectifier), the approximate ripple factor is obtained by setting $n = p$ in (8.6) and taking the ratio of (8.6) to the average value \bar{E}_z of the output voltage. Then

$$\text{R.F.} \doteq \frac{E_{p2}}{\bar{E}_z} = \frac{E_p/\bar{E}_z}{p^4\omega^4 L_1 L_2 C_1 C_2} = \frac{k_p}{p^4\omega^4 L_1 L_2 C_1 C_2} \quad (8.7)$$

where k_p is the first or largest value of k_n (corresponding to $n = p$) given in Table 6.1 for type of rectifier circuit employed.

¹ In some cases the factor $n^2\omega^2 L_1 C_1$ is not large compared with unity so that the simplification used in obtaining 8.2 and 8.3 is not justified. In this case the exact value $(n^2\omega^2 L_1 C_1 - 1)$ should be used in the denominator of Eq. 8.3.

9. Critical Inductance.—The effect of varying the inductance of the first choke of a two-section choke-input filter while the second choke and the capacitors are kept constant is shown¹ by Fig. 9.1. As the inductance L_1 is reduced from a large value, the output voltage of the filter remains constant at the average value of the rectified wave until a critical value L_c is reached. For an inductance less than L_c the first choke ceases to function satisfactorily, and the capacitor C_1 , Fig. 8.1, charges to a higher voltage. The voltage across C_1 approaches the peak value of the alternating voltage of the rectifier transformer as a limit.

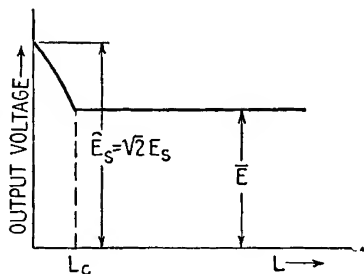


FIG. 9.1.—Effect of choke inductance L on average output voltage of choke-input filter; \bar{E} = average of full-wave rectifier sinusoid of amplitude \hat{E}_s .

The first choke of a choke-input filter should be large enough so that the pulsating output current of the rectifier never goes to zero. An approximation of L_1 sufficiently accurate for most

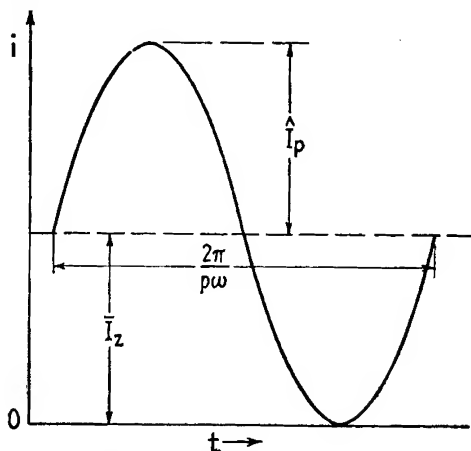


FIG. 9.2.—Limiting case of input current to a choke-input filter.

engineering purposes can be made by assuming that the pulsating component of rectified current is due entirely to the lowest har-

¹ F. S. DELLENBAUGH, JR., and R. S. QUIMBY, *QST*, February, 1932, p. 14, March, 1932, p. 27, April, 1932, p. 33; H. J. REICH, "Theory and Application of Electron Tubes," 1st ed., p. 547, 2nd ed., p. 585, McGraw-Hill Book Company, Inc., 1939, 1944; W. P. OVERBECK, *Proc. I.R.E.*, **27**, 655, 1939.

monic p present. The limiting case is shown in Fig. 9.2, where \hat{I}_p is the peak value of the lowest harmonic present. By (8.3),

$$\hat{I}_p = \frac{\sqrt{2} E_p}{p\omega L_c} = \bar{I}_z = \frac{\bar{E}_z}{R} \quad (9.1)$$

where R is load plus choke resistance in ohms. The value of E_p is given in Table 6.1 as $k_p \bar{E}_z$ so that (9.1) can be written as

$$L_c = \frac{\sqrt{2} k_p R}{p\omega} \quad \text{henrys} \quad (9.2)$$

In general, for polyphase rectifiers,¹ the critical inductance is given² by

$$L_c = \frac{R}{\pi f p (p^2 - 1)} \quad \text{henrys} \quad (9.3)$$

where f is the fundamental frequency of the power source. For the full-wave rectifier on 60 cycles, (9.2) gives

$$L_c = \frac{R}{1,132} \quad (9.4)$$

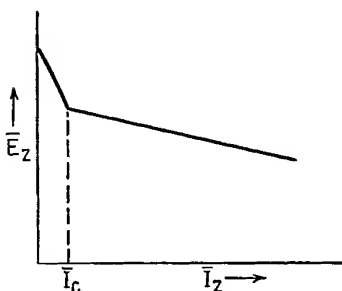


FIG. 9.3.—Voltage-regulation curve of a choke-input power supply.

As a matter of conservative design it is best to take $L_1 = 2L_c$.

For a single-phase half-wave rectifier, $p = 1$, and L_c is infinite. This means that ideally there is no value of inductance which will prevent the tube current from be-

coming zero during some part of the cycle (see Fig. 2.9a).

Varying Loads.—When the load resistance is varied, the behavior of the power supply will be erratic unless (9.1) is satisfied for all values of R . The effect on the output voltage of varying the load current of a choke-input filter is shown in Fig. 9.3. As the current is decreased from its normal value, the output voltage rises owing to a decrease in the voltage drops in the resistance of the chokes and transformer. When the d-c load current is decreased to the value \bar{I}_c at which the first choke ceases to filter properly, the voltage rises steeply to the peak value of the transformer voltage per anode.

The rapid increase in the voltage when \bar{I}_z is less than \bar{I}_c may be

¹ COBINE, *op. cit.*, Chap. XI.

² MILLMAN and SEELY, *op. cit.*, p. 468.

prevented by connecting a resistance permanently across the filter output so that even when the useful load is zero the current through the filter due to this resistance will be greater than \bar{I}_c . Such a resistance is called a "bleeder" resistance. The bleeder also serves to discharge the filter capacitors when the load resistance and supply voltage are removed. Another solution is to use a choke coil whose inductance is very high when the d-c current is low but is lower at higher currents. Such a choke has an iron core that becomes saturated when the d-c current is high but operates on the lower part of the BH curve at low currents and thus presents a high effective inductance. Such a choke is called a "swinging" choke;¹ manufacturers specify the upper and lower values of the inductance.

10. Design of a Power Supply.—As an example of the steps taken in the design of a power supply involving a choke-input filter, assume that a 60-cycle 110-volt power source is to be rectified to give 2,500 volts d-c and 400 ma with a ripple not to exceed 1 per cent ($R.F. = 0.01$). A full-wave rectifier circuit and a one-section filter will be used.

In order to make allowance for resistance drops it is best to start with the output of the filter and work back to the transformer. At rated load current the load resistance is $R = 2,500/0.4 = 6,250$ ohms. The critical inductance for full-wave 60-cycle rectification is $L_c = R/1,132 = 6,250/1,132 = 5.55$ henrys. For conservative design, $L_1 = 2L_c = 11.1$ henrys. A 10-henry choke ordinarily will be available, which would be satisfactory in most cases, although a larger choke may be used. If the load current is to vary over a considerable range a 10- to 25-henry swinging choke may be used.

If the additional protection of a bleeder is desired, its resistance may be determined from (9.4) and the maximum resistance of the bleeder as $R = 25 \cdot 1,132 = 28,300$ ohms. This gives a bleeder current of $2,500/28,300 = 0.088$ ampere with a power loss of $(2,500)^2/28,300 = 219$ watts. If this protection is required, the rectifier must supply a current of 0.488 ampere instead of the original 0.4 ampere. A higher resistance bleeder may be used if some voltage rise is permissible.

The chokes should be placed in the ungrounded lead when possible.² This location prevents the chokes from being by-passed by the capacitance of the secondary of the transformer to ground. Magnetic coupling between the transformer and the chokes should

¹ R. M. HANSON, *Electronics*, June, 1943, p. 112.

² F. E. TERMAN and S. B. PICKLES, *Proc. I.R.E.*, **22**, 1040, 1934.

be avoided. The chokes should be insulated for peak value of the voltage applied to the anodes of the rectifier tube.

The first capacitor C_1 , Fig. 8.1, may be found from the ripple limitations. The rms value of the second harmonic present in the voltage wave for a full-wave rectifier is $0.47\bar{E}_z$, Table 6.1. By (8.3) the voltage developed across C_1 due to this input harmonic alone is

$$E_{21} = \frac{E_2}{2^2(2\pi 60)^2 L_1 C_1}$$

so that for a ripple factor of 0.01

$$\text{R.F.} = \frac{E_{21}}{\bar{E}_z} = 0.01 = \frac{0.47}{4(377)^2 \cdot 11.1 C_1}$$

Solving for C_1 gives a value of 7.4 μf , so that an 8- μf capacitor would be used to accomplish the desired smoothing in one stage. As a precaution, a check should be made to determine whether the section $L_1 C_1$ will resonate at the lowest harmonic frequency present. In this case the resonant frequency is 17.7 cps. A second stage duplicating the first would produce a very smooth output voltage wave having R.F. = 0.00018.

Having selected the first choke, which may be assumed to have a d-c resistance of 100 ohms for example, the output voltage of the rectifier is found by adding the normal d-c voltage drop in the choke to the load voltage, as

$$\bar{E}_z = 2,500 + 0.4 \cdot 100 = 2,540 \text{ volts}$$

(If the bleeder is used, the bleeder current would increase the drop in the choke coil to $0.488 \cdot 100$, or 49 volts, so that the output voltage of the rectifier would have to be increased to 2,549 volts.) The transformer voltage per anode E_s of the full-wave rectifier is found from the constant E_s/\bar{E}_z of Table 7.1. The tube drop of 10 volts is added to the d-c voltage so that

$$E_s = \frac{2,540 + 10}{0.9} = 2,835 \text{ volts rms}$$

This is the secondary voltage on each side of the center tap. The capacitors must withstand the peak of this voltage, $2,835\sqrt{2}$, or 4,010 volts. If it is necessary to use capacitors in series to withstand this voltage, they may be shunted by resistors to equalize the voltage properly.

The volt-ampere rating of each section of the transformer secondary winding is by (7.1) and Table 7.1

$$VA(\text{sec}) = \frac{\bar{E}_s \bar{I}_s}{\text{U.F.}(\text{sec})} = \frac{2,540 \cdot 0.4/2}{0.637} = 797$$

The total secondary rating is $2 \cdot 796 = 1,592$ volt-amperes. The rating of the primary winding is

$$VA(\text{pri}) = \frac{\bar{E}_s \bar{I}_s}{\text{U.F.}(\text{pri})} = \frac{2,540 \cdot 0.4}{0.9} = 1,129$$

These ratings must be used if the transformer is to be designed. If a transformer designed for the usual a-c purposes is to be used, an average utility factor of $(0.637 + 0.9)/2 = 0.768$ can be used. This would call for a transformer of $1,000/0.768 = 1,300$ volt-amperes, which would not overheat when supplying a 1,000-watt rectified load. Since the primary voltage is known, this completes the information necessary for the selection of a suitable transformer.

In order to select a rectifier tube the average and peak tube currents must be determined, Table 7.1. The average current per tube is one-half the average load current, or 0.2 ampere. With such a large choke in the filter the peak tube current is approximately equal to the load current, or 0.4 ampere. Since capacitors are present, the tubes must withstand an inverse voltage of twice the peak applied voltage per anode when the load is reduced to a low value; therefore,

$$\hat{E}_{inv} = 2.82E_s = 2.82 \cdot 2,835 = 7,990 \text{ volts}$$

This completes the information necessary for the selection of a suitable rectifier tube. An 866A tube would be satisfactory for this application. This is a mercury-filled diode having an average current rating of 0.25 ampere, a peak current of 1.10 amperes, a peak inverse voltage of 10,000 volts (at 60°C), and a tube drop of 10 volts.

Filament Transformers.—For small sizes the cathodes of rectifier tubes can be heated by auxiliary windings on the power transformer. Larger sizes of tubes should have separate filament transformers. It is best to center-tap the filament transformers in order to equalize the flow of the rectified current in the winding, especially if the d-c current is of the same order of magnitude as the heating current for the filament emitter.

11. Capacitor-input Filters.—Capacitor-input filters,¹ Fig. 11.1, are used where a higher voltage is desired than can be obtained by a choke-input filter. The choke-input filter has an output voltage that approximates the average value of the rectified voltage, whereas the capacitor-input filter has an output voltage that approximates the peak value of the rectified voltage. However, the voltage

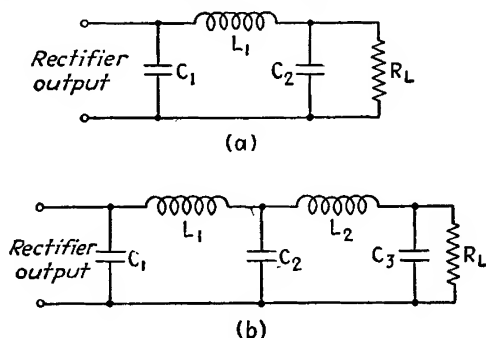


FIG. 11.1.—Capacitor-input filters: (a) one-section and (b) two-section.

regulation of the capacitor-input filter is not as good as that of a choke-input filter, the voltage dropping off rapidly as the load current is increased. Another objectionable feature of the capacitor-input filter is the high ratio of peak current to average current per tube. This type of filter can be roughly analyzed by using the ripple voltage indicated by the curves for the RC load, Secs. 2 and 3. The saw-tooth voltage from the first capacitor is applied to the

first choke and second capacitor; these filter out the ripple, acting in the same manner as one section of a choke-input filter. This results in a smoother output voltage than is possible with a choke-input filter.

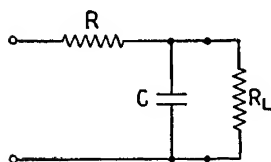


FIG. 12.1.—Resistance-capacitance filter.

12. Resistance-Capacitance Filter.—

In many cases a simple RC filter such as is shown in Fig. 12.1 will provide adequate smoothing of the rectified voltage. This filter is also used as a decoupling element between the plate supplies of multiple-stage amplifiers. An approximate expression for the ripple factor for the RC filter may be found as follows: The impedance of the capacitor to the lowest harmonic present in the output voltage of the rectifier will be assumed very low compared with R and R_L . Then the rms value

¹ Terman, *op. cit.*, p. 491.

of current flowing through R and C due to the harmonic voltage E_n of angular frequency $n\omega$ is

$$I_n = \frac{E_n}{R}$$

The ripple voltage developed across C due to the passage of the current I_n is

$$E_{nc} = \frac{I_n}{n\omega C} = \frac{E_n}{n\omega CR}$$

Then the ripple factor is

$$\text{R.F.} = \frac{k_n}{n\omega CR}$$

where $k_n = E_n/\bar{E}_z$ as given in Table 6.1 for the lowest harmonic n_1 present in the output voltage. The ripple can be further reduced by using an input capacitor to make a Π section.

13. Voltage-regulated Power Supplies.¹—Many electronic devices require a source of direct current whose output voltage is constant with varying load and is free from cyclic variations or ripples. The output voltage of a rectifier and filter combination decreases as the load current increases, owing to resistance drops in the filter elements and in the rectifier transformer. Variations in the a-c supply voltage also cause the output of a rectifier and filter to vary. The simplest type of filter, a large capacitor supplied directly from the rectifier, has very poor regulation and pronounced ripples.

One solution of the stabilization of output voltage consists in placing a variable impedance device, or stabilizer, between the filter output and the load, Fig. 13.1. The impedance of the stabilizer is controlled by the output voltage so as automatically to hold that voltage constant. Such a device can maintain a constant output voltage \bar{E}_o at the load, by absorbing a varying amount of the output voltage of the filter or input voltage \bar{E}_i to the stabilizer, as shown by the shaded area of Fig. 13.1a. This system can be represented schematically, Fig. 13.1b, where \bar{E}_i' is the equivalent no-load output voltage² of the rectifier, R_r is the equivalent internal resistance of the rectifier and filter, \bar{E}_i is the input voltage of the stabilizer, \bar{E}_o is the output voltage of the stabilizer, R is the load resistance, e is a small variational or test voltage, S_0 is the stabilization ratio, defined by $1/S_0 = \partial\bar{E}_o/\partial\bar{E}_i$ ($e = \text{const.}$), and R_0 is the

¹ F. V. HUNT and R. W. HICKMAN, *Rev. Sci. Instruments*, **10**, 6, 1939.

² See Appendix C.

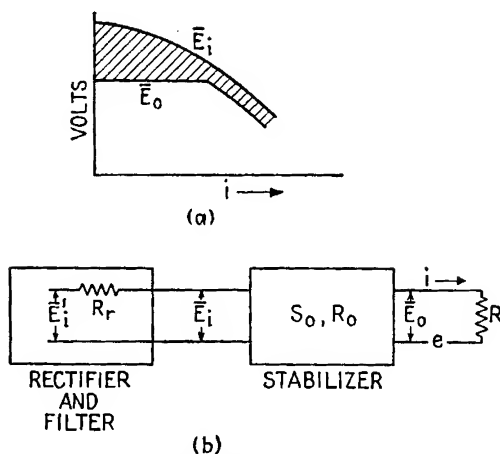


FIG. 13.1.—(a) Input and output voltages of the stabilizer; (b) equivalent circuit of a stabilized power supply.

internal resistance of the stabilizer, defined by $1/(R_o + R) = \partial i / \partial e$ ($\bar{E}_i = \text{const.}$). A low internal resistance of the stabilizer is desirable when several tubes are supplied from the same source in order to reduce the coupling between tubes.

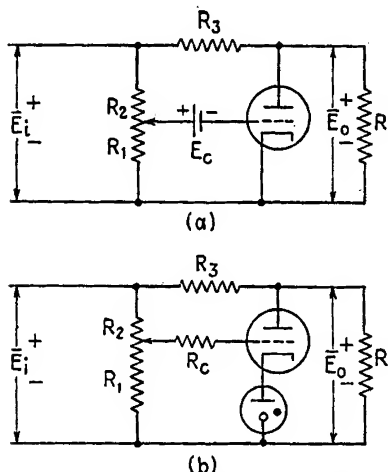


FIG. 13.2.—Transconductance stabilizer.

Transconductance Type of Stabilizer (g_m Circuit).—This circuit is based on the bridge circuit used in measuring $g_m = \partial i_p / \partial e_g$ ($e_p = \text{const.}$). The bridge circuit, Fig. 13.2, produces such variations of grid voltage and plate current that the plate voltage of the tube is held constant. In this circuit the load resistance R replaces the phones of the normal bridge circuit. The grid bias is varied by changes in \bar{E}_i , and thus the current in R_3 is caused to vary. Suppose \bar{E}_i

momentarily increases; this makes the grid more positive relative to the cathode, causing the current through the tube, which shunts the load resistance, to increase. This increased current flowing through R_3 causes an increased voltage drop in R_3 , which, with proper adjustment, can be made to counteract exactly the

original increase in \bar{E}_i so that the output voltage \bar{E}_o remains constant. The adjustment for balance is given by

$$R_3 = \frac{R_1 + R_2}{g_m R_1}$$

In this stabilizer

$$S_0 = \infty$$

and

$$R_0 = \frac{r_p}{1 + \mu \left(\frac{R_1}{R_1 + R_2} \right)}$$

The transconductance stabilizer is good primarily for constant load since the bridge will not remain balanced for large variations of \bar{E}_i and current. The entire output voltage is across the tube so that the circuit cannot be used for high voltages unless large tubes are used. The common terminal is negative, which is the usual ground polarity. For low internal resistance a low- μ triode having a large g_m should be used. The bias battery can be eliminated by the use of a glow tube operating on the constant-voltage portion of its characteristic, Fig. 13.2b. In this circuit the glow tube maintains the cathode of the stabilizer tube at a constant voltage above the negative side. The grid therefore can be made negative relative to the cathode by decreasing R_1 relative to R_2 so that the voltage across R_1 is less than that across the glow tube.

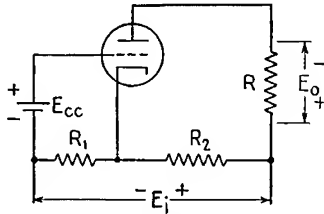


FIG. 13.3.—Amplification-factor stabilizer.

Amplification-factor Bridge Type of Stabilizer.—This stabilizer is based on the circuit used to determine $\mu = \partial e_p / \partial e_g$ ($i_p = \text{const.}$) and introduces voltage variations in both grid and plate circuits so that i_b is held constant. In Fig. 13.3 the load resistance R replaces the phones of the bridge circuit. At balance there is no variational voltage across R_L since the plate current is held constant. The condition of balance occurs for $\mu = R_2 / R_1$. The action of the stabilizer may be understood by considering the path of load current flowing from the positive terminal through the load resistance R , the tube, and R_1 to the negative input terminal. A momentary increase in \bar{E}_i will cause the cathode to become more positive so that the grid, which has a fixed bias, becomes more negative relative

to the cathode. This causes the tube resistance to be increased, yielding a higher tube drop for a constant current so that the output voltage \bar{E}_o is held constant. A decrease in R , causing a momentary increase in plate current, will also make the cathode more positive because the increased drop in R_1 and the negative grid bias tend to counteract the increase in plate current. This stabilizer has $S_0 = \infty$ and $R_0 = r_p + R_2$. The stabilizer does not have a low internal resistance. The positive leads of \bar{E}_i and \bar{E}_o are common, which is sometimes an advantage. Since the tube does not have to withstand the entire input voltage, a receiving-type tube can be used

for voltages as high as 2,000 to 3,000 volts.

Degenerative Type of Stabilizer.—Neither of the previous stabilizers yields satisfactory voltage regulations when the changes in load current cover a considerable range. The degenerative amplifier type of stabilizer functions for changes in both input and output voltages. In this circuit the load resistance is placed between the cathode and the negative lead, Fig. 13.4a. A momentary increase

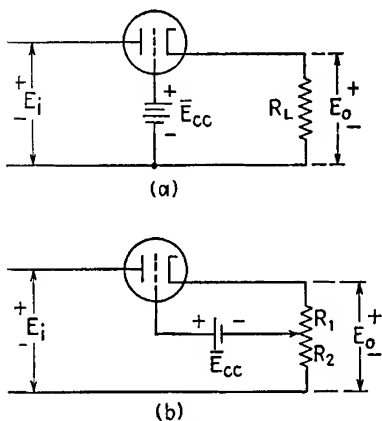


FIG. 13.4.—Degenerative stabilizer.

in \bar{E}_o causes the cathode to become more positive relative to the grid so that the resistance of the tube is increased, *i.e.*, the tube drop increases, so as to prevent the increase in load voltage, thus holding the output voltage constant. In this circuit

$$S_0 = 1 + \mu + \frac{r_p}{R_L}$$

and

$$R_0 = \frac{r_p}{S_0 - \frac{r_p}{R_L}}$$

A more practical circuit is shown in Fig. 13.4b. In this circuit the grid-bias battery need not be of the same order of voltage as \bar{E}_o , and the value of \bar{E}_o is continuously variable by the tap between R_1 and R_2 . With the modification the stabilization ratio is

$$S_0 = 1 + \frac{\mu R_1}{R_1 + R_2} + \frac{r_p}{R_L}$$

while the expression for R_0 , which involves the new S_0 , remains unchanged.

Combination Stabilizer.—The necessity for a bias battery can be removed by the use of a glow tube, and an amplifier can be added to improve the flexibility and stabilization with the degenerative type of stabilizer, Fig. 13.5. The cathode of T_2 is at a positive voltage determined by the operating voltage of the glow tube T_3 .

The grid bias of T_2 is the difference between the glow voltage of T_3 and the voltage drop across R_3 . If \bar{E}_o increases, the grid of T_2 is made more positive relative to its cathode. This causes the resistance of T_2 to decrease and its plate current to increase. This plate current flows through R_2 and causes an increased potential

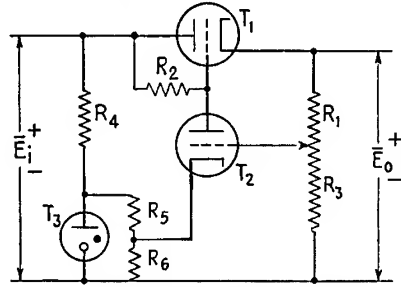


FIG. 13.5.—Combination stabilizer.

difference between the plate and grid of T_1 and drives the grid of T_1 more negative. This increases the resistance of T_1 , producing a higher plate voltage that counteracts the increase in \bar{E}_o by subtracting a greater voltage from the input. The stabilization ratio is

$$S_0 = \frac{1 + \mu_1 \left(1 + \frac{G_2 R_3}{R_1 + R_3} \right) + \frac{r_{p1}}{R}}{1 + \mu_1 \left(1 - \frac{G_2}{\mu_2} \right)}$$

where G_2 is the gain of the amplifier stage,

$$G_2 = \frac{\mu_2 R_2}{r_{p2} + R_2 + r(1 + \mu_2)}$$

where r is the equivalent resistance of the glow tube at the current corresponding to normal operation of the stabilizer. The load R is the combined effect of the external load and the internal resistances R_1 and R_3 . The internal resistance of the stabilizer is

$$R_0 = \frac{r_{p1}}{S_0 - \frac{r_{p1}}{R}}$$

The triode amplifier tube T_2 , Fig. 13.5, may be replaced by a pentode¹ with a marked increase in stability. The same effect can be obtained by using two triodes in series,² Fig. 13.6, as a part of a complete power supply. The effective amplification factor of tubes T_2 and T_3 in series is

$$\mu = \mu_2 + \mu_3 + \mu_2\mu_3$$

while the plate resistance of the combination is

$$r = r_{p2}(1 + \mu_3) + r_{p3}$$

For a 6C8G tube in this application the effective amplification may be 1,300 with a plate resistance of 3 to 5 megohms. Tube T_1 may

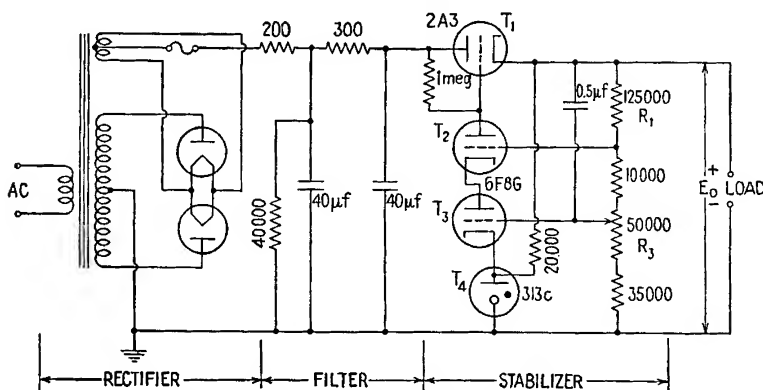


FIG. 13.6.—Complete voltage-regulated power supply.

be a 2A3, T_2 and T_3 a 6C8G or 6F8G double triode. The current rating may be increased by using two or more 2A3 power-amplifier tubes in parallel. A capacitor is placed between the cathode of T_1 and the grid of T_3 greatly to increase the effectiveness of the stabilizer when rapid fluctuations and ripples are present. This capacitor effectively short-circuits R_1 and applies the entire ripple voltage across R_3 . This circuit may be several times as effective in suppressing ripples as in correcting for slow variations of voltage. The stabilization ratio may be from 50 to ∞ , and the internal resistance is 1 or 2 ohms. Since a very effective stabilizer is used in the power supply of Fig. 13.6, a very simple CRC filter suffices. A 200-ohm resistor placed in the cathode lead protects the rectifier tube from excessively high peak currents when the alternating cur-

¹ A. B. BERESKIN, *Proc. I.R.E.*, **31**, 47, 1943.

² "Cascode"; see Hunt and Hickman, *loc. cit.*

rent is switched on with the capacitors uncharged. Under some conditions a stabilizer circuit may develop switching transients, which appear in the output voltage. These may be suppressed in most cases by placing a capacitor across the output terminals.

Glow Lamps.—The characteristics of glow lamps such as the WE313C, 874, 0C3/VR105, and 0D3/VR150 are similar to those of the lamp shown in Fig. 13.7. The most desirable point of operation, viz., for the normal adjustment of R_4 , Fig. 13.5, would be in the center of the bc region. The negative region ab is quite variable. In some tubes, as the current is varied, the region bc is unstable owing to changes in the active area of the cathode. In this case it has been found desirable to design the stabilizer for operation in the cd region, where the tube is equivalent to a battery whose voltage is

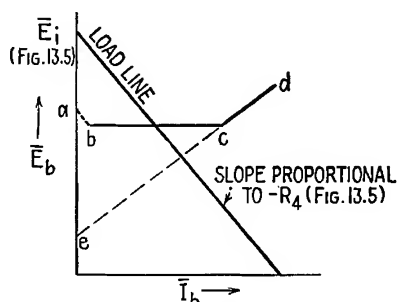


FIG. 13.7.—Characteristic of glow tubes, with d-c load line corresponding to operation as in Fig. 13.5.

the voltage axis intercept e of the characteristic cd extended and whose resistance is given by the slope of the characteristic cd .

Internal Factors Affecting Stability.—The d-c stabilizing circuit of Fig. 13.6 is similar to a high-gain d-c amplifier; therefore, spurious d-c voltages introduced in the circuit will have serious effects. One such source is in the equivalent emf existing in the grid-cathode circuit of the amplifier tube, representing the combined effects of contact potential, initial velocity of emission, etc. This emf may change by as much as 0.1 volt for a 10 per cent change in heater current. A cascode stabilizer to deliver $1,000 \pm 1$ volts for a ± 5 per cent line-voltage variation would normally require $S_0 = 50$. However, if the grid voltage of T_3 is 0.25 of \bar{E}_o , the spurious emf would account for 80 per cent of the allowed variation so that an $S_0 = 250$ would be required. Therefore, a saturable-core type of line-voltage regulator should be used, or the heater

circuit of the first amplifier of the cascode stabilizer may be stabilized by an iron-wire ballast tube.

Current Regulator.—It is sometimes essential that a constant current be maintained in a circuit, independent of voltage over a

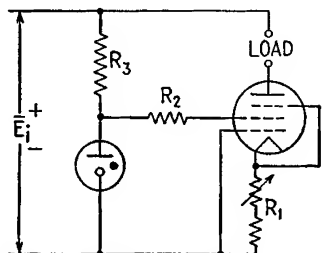


FIG. 13.8.—Pentode current stabilizer.

wide range, as in the exciting winding of an electromagnet for special applications. This may be accomplished by the stabilizer, Fig. 13.8. The circuit is adjusted to operate on the portion of the pentode characteristic for which plate current is little affected by plate voltage.

CHAPTER XVIII

SIGNAL ANALYSIS

1. Statement of the Problem.—In this chapter the word *signal* is used for “a voltage varying nonperiodically with time.” The purpose in sending a signal is to transmit intelligence; to that effect any physical variable (such as light intensity, sound pressure, etc.)



FIG. 1.1.—Interrupted constant voltage (telegraphic transmission).

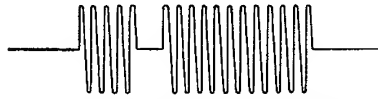


FIG. 1.2.—Interrupted sinusoidal voltage (continuous-wave telegraphy).

may be used instead of voltage, which is given as an example. A constant voltage, or a periodic voltage in the strict sense (periodic for all times), cannot be used for transmitting intelligence. Some examples of varying voltage, that can be and have been effectively used for the transmission of intelligence are given in Figs. 1.1 to 1.3.



FIG. 1.3.—Modulated alternating voltage (short a-c pulse).

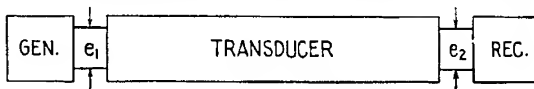


FIG. 1.4.—Signal transmission.

The signal originates at a *generator* and is intended for a *receiver*. The *transmission path* between generator and receiver may be thought of as an electrical network, Fig. 1.4, having two pairs of terminals, with one pair, the *input* terminals, connected to the

generator and with the other pair, the *output* terminals, connected to the receiver. In transmission theory, such a network is called a *transducer*. More often the transmission path is analyzed as a number of transducers connected in tandem. It will be assumed that the transducer is a *linear* network, *i.e.*, that its "constants" R, L, C, \dots are independent of the voltages and currents present in the network, Sec. 23.

Let $e_1(t)$ be the voltage (signal) applied at the input terminals of the transducer. This voltage comes out of the transducer as $e_2(t)$, the difference of potential between the output terminals. As a rule, it is desirable that $e_2(t)$ have exactly the same shape as $e_1(t)$, which, however, is usually not the case. The signal is *distorted*, the transducer and load having caused the distortion. The problem is to find the shape of $e_2(t)$, when the input signal $e_1(t)$, the transmitter, and the load are known. From this information, the causes of distortion can be determined. Sometimes it is desired to find the current $i_2(t)$ flowing into the load; this problem is fundamentally the same as finding $e_2(t)$ and is solved by the same method, replacing "voltage ratio" by "transfer impedance."

There are two essentially different methods for finding $e_2(t)$. The first method consists in solving the system of differential equations of the network, the right-hand member being $e_1(t)$ for one equation and zero for all the others. This method is practical only when the number of meshes in the network is one, two, or three, but then it gives the most accurate and complete answer. Examples of this method are given in Chap. VI.

The second method makes use of a generalization of the Fourier analysis of a periodic signal. It defines the transducer by way of its attenuation and phase-lag characteristics and analyzes the non-periodic signal by means of two continuous frequency spectra, *viz.*, amplitude and phase. The second method is used in this chapter.

FOURIER ANALYSIS OF SYMMETRICAL PULSE

2. Transmission of Nonperiodic Voltage.—It is shown in Chap. IX how to use the Fourier development of a periodic input voltage e_1 and the frequency characteristics of a network to obtain the periodic output voltage e_2 . However, a signal intended to transmit intelligence must be nonperiodic, Sec. 1. This will necessitate a generalization of the treatment in Chap. IX. In the case of a periodic input voltage, frequencies other than multiples of the

fundamental frequency f were not used at all. It will now be shown that a nonperiodic signal (a single Morse dot, for example) can be analyzed into a sum of components of every frequency. This will necessitate the use of integral notation, and the formulas that will be obtained are called *Fourier integrals*.

3. Rectangular Pulse.—The first signal that will be analyzed is a *single* Morse “dot” of amplitude E volts, lasting $2t_1$ sec, from $t = -t_1$ to $t = +t_1$, Fig. 3.1. As a first step in the analysis, let the dot be repeated every T sec (both before and after $t = 0$). A periodic voltage, Fig. 3.2, results that can be expressed by a Fourier series. If T is increased without limit, the dot centered on $t = 0$ will become, for all practical purposes, an isolated rectangular pulse and will be found expressed as a Fourier integral.

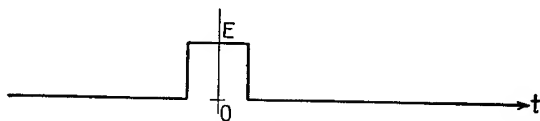


FIG. 3.1.—Single Morse dot.

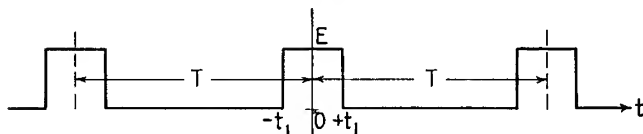


FIG. 3.2.—Periodic succession of Morse dots.

The amplitude b_n of the n th harmonic component $b_n \cos n(2\pi/T)t$ of a symmetrical periodic voltage $C(t)$ is twice the average of $C(t) \cos n(2\pi/T)t$ over one period, or, in integral form,

$$b_n = \frac{2}{T} \int_{-T/2}^{+T/2} C(t) \cos \left(n \frac{2\pi t}{T} \right) dt \quad (3.1)$$

as in (6.6), Chap. IX. In the present case,

$$b_n = \frac{4E}{T} \int_0^{t_1} \cos \left(n \frac{2\pi}{T} t \right) dt \quad (3.2)$$

$$= \frac{4Et_1}{T} \frac{\sin \frac{2\pi n t_1}{T}}{2\pi n t_1 / T} \quad (n = 1, 2, 3, \dots) \quad (3.3)$$

In increasing T , assume first that $T = 4t_1$ (twice the duration of the pulse), Fig. 3.3. The alternating part of the voltage, apart from the constant component, is a square wave. From (3.3),

$$b_n = E \frac{\sin n \frac{\pi}{2}}{n \frac{\pi}{2}} \quad (3.4)$$

whence

$$b_1 = \frac{2E}{\pi}, \quad b_3 = -\frac{b_1}{3}, \quad b_5 = \frac{b_1}{5}, \quad \dots \quad (3.5)$$

The amplitudes b_1, b_3, b_5, \dots are represented graphically in Fig 3.4.

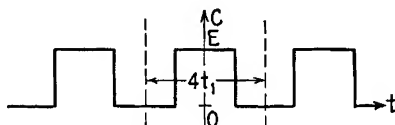


FIG. 3.3.—Case of $T' = 4t_1$.

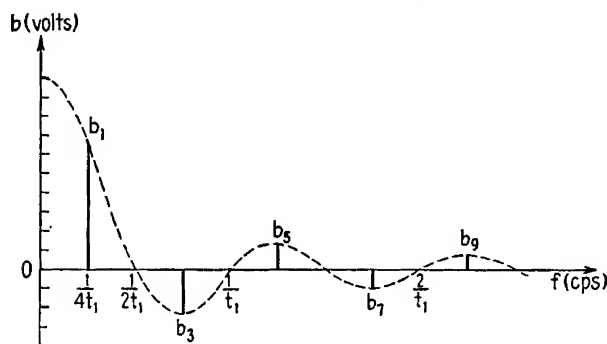


FIG. 3.4.—Fourier spectrum of the periodic voltage of Fig. 3.3.

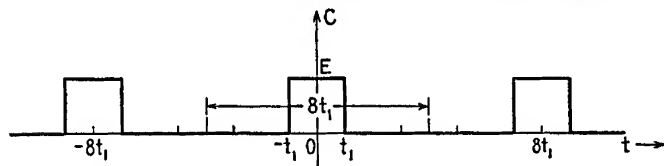


FIG. 3.5.—Case of $T' = 8t_1$.

If the period of the periodic voltage is doubled, the duration $2t_1$ of the rectangular pulse, Fig. 3.5, being kept constant, then $T' = 2T = 8t_1$ and, from (3.3),

$$b_n' = \frac{E}{2} \frac{\sin n \frac{\pi}{4}}{n \frac{\pi}{4}} \quad (n = 1, 2, 3, \dots) \quad (3.6)$$

whence

$$b_1' = \frac{E \sin(\pi/4)}{2 \frac{\pi}{4}} \quad b_2' = \frac{E \sin(\pi/2)}{2 \frac{\pi}{2}} \quad \dots$$

or

$$b_1' = \frac{\sqrt{2} E}{\pi} \quad b_2' = \frac{E}{\pi} \quad b_3' = \frac{\sqrt{2} E}{3\pi} \quad b_4' = 0 \quad \dots \quad (3.7)$$

Thus the fundamental frequency for $T' = 8t_1$ is one-half the value for $T = 4t_1$, and the corresponding Fourier spectrum, Fig. 3.6, has twice as many bars as that of Fig. 3.4.

If the period of $C(t)$ is continually doubled, the fundamental frequency $1/T$ at each successive stage will be reduced to one-half its preceding value. The interval $\Delta f = 1/T$ between two adjacent component frequencies is also halved so that each new Fourier spectrum contains twice as many bars as the preceding one.

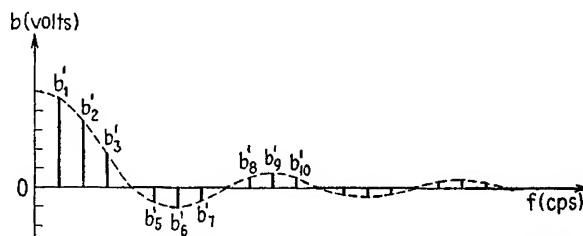


FIG. 3.6.—Fourier spectrum of the periodic voltage of Fig. 3.5.

When T becomes very large, Δf becomes very small. The component angular frequencies, $\omega_n = 2\pi n/T = 2\pi n(\Delta f)$, get closer and closer together and in the limit merge into one continuous variable ω . The Fourier spectrum, composed of vertical bars closer and closer together, becomes in the limit a continuous frequency spectrum.

The height of the bars, which represents the amplitude of the successive harmonic components, can be written, from (3.3), as

$$b_n = 4Et_1(\Delta f) \frac{\sin 2\pi t_1 n(\Delta f)}{2\pi t_1 n(\Delta f)} \quad (3.8)$$

since $\Delta f = 1/T$.

As Δf approaches zero, b_n also approaches zero; but the ratio $b_n/\Delta f = b_n T$ does not. Let $B(\omega)$ be the limit of the ratio,

$$\frac{b_n}{2\pi(\Delta f)} = \frac{b_n}{\Delta\omega}$$

which will be a number of volt-seconds (or webers) depending upon the angular frequency ω . Since in the limit the quantity $2\pi n(\Delta f) = \omega_n$ becomes the continuous variable ω , $b_n/2\pi(\Delta f)$, from (3.8), becomes

$$B(\omega) = \frac{2Et_1}{\pi} \frac{\sin t_1\omega}{t_1\omega} \quad (3.9)$$

$B(\omega)$ as function of frequency is shown in Fig. 3.7. For $t_1\omega$ greater than π , the curve is similar to that for a slowly damped sinusoid. As $t_1\omega$ approaches zero, the ratio $(\sin t_1\omega)/t_1\omega$ approaches unity and $B(\omega)$ approaches $2Et_1/\pi$.

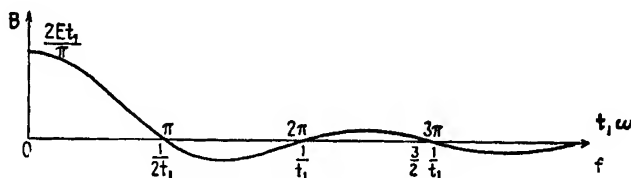


FIG 3.7.— $B(\omega)$ as function of f .

Mathematically, the periodic square wave for $T = 4t_1$ is equal to the sum of the Fourier series

$$\frac{4E}{\pi} \left(\cos \frac{\pi t}{2t_1} - \frac{1}{3} \cos \frac{3\pi t}{2t_1} + \frac{1}{5} \cos \frac{5\pi t}{2t_1} - \dots \right)$$

As T is increased without limit, the periodic wave becomes the sum of harmonic components more and more closely packed within a given frequency band; for T equal to infinity the Fourier spectrum becomes a continuous frequency spectrum, and the Fourier series becomes the Fourier integral

$$\int_0^\infty B(\omega) \cos \omega t d\omega \quad (3.10)$$

where $B(\omega)$ is given by (3.9).

The unique rectangular pulse thus is represented as a sum of harmonic components. The amplitude of the harmonic component of angular frequency ω is $B(\omega) d\omega$. It is infinitely small, which is consistent with the fact that all frequencies are now component frequencies. The curve $B(\omega)$ thus represents the relative importance, or *weight*, pertaining to the individual harmonic component $B(\omega) \cos \omega t d\omega$ as ω increases from zero to infinity. The infinitely small amplitude $B(\omega) d\omega$ means that the fraction of the energy of the rectangular pulse involved in an infinitely narrow frequency

band ω to $\omega + \Delta\omega$ is infinitely small; if the band width is finite, the corresponding fraction of the energy is finite.

4. Symmetrical Voltage Pulse.—The reasoning that has just been applied to the rectangular pulse can be applied to any pulse $C(t)$ of symmetrical shape (as defined in Chap. IX). Such a unique pulse is first repeated at equal intervals of time T and thus gives rise to a periodic function $C_1(t)$, which can be represented by a Fourier series. Because the original pulse is symmetrical, the series will contain cosines only and

$$C_1(t) = b_0 + b_1 \cos \frac{2\pi}{T} t + \cdots + b_n \cos n \frac{2\pi}{T} t + \cdots \quad (4.1)$$

The period T is then increased without limit. The Fourier spectrum becomes a continuous frequency spectrum, and the Fourier series (4.1) becomes the Fourier integral

$$C(t) = \int_0^\infty B(\omega) \cos \omega t d\omega \quad (4.2)$$

The function of frequency $B(\omega)$, such that $B(\omega) d\omega$ is the amplitude of the component $\cos \omega t$, is obtained as follows: From (3.1),

$$b_n = \frac{2}{T} \int_{-T/2}^{+T/2} C(t) \cos \left(n \frac{2\pi}{T} t \right) dt \quad (4.3)$$

and since $C(t)$ is symmetrical,

$$b_n = \frac{2}{\pi} \frac{2\pi}{T} \int_0^{T/2} C(t) \cos (\omega_n t) dt \quad (4.4)$$

As T increases indefinitely, in the limit ω_n becomes the continuous variable ω , and $2\pi/T = 2\pi(\Delta f)$ becomes $d\omega$. The amplitude b_n becomes $B(\omega) d\omega$, so that in the limit

$$B(\omega) d\omega = \frac{2}{\pi} d\omega \int_0^\infty C(t) \cos \omega t dt$$

and

$$B(\omega) = \frac{2}{\pi} \int_0^\infty C(t) \cos \omega t dt \quad (4.5)$$

Thus, when the symmetrical voltage pulse $C(t)$ is known as a function of time, a function of frequency $B(\omega)$ can be derived from it, by formula (4.5). Conversely, when a function of frequency $B(\omega)$ is given, a symmetrical function of time $C(t)$ can be derived

from it, by formula (4.2). To each symmetrical voltage pulse corresponds a single continuous frequency spectrum, and vice versa; thus, it is indifferent which is given. However, the spectrum lends itself readily to calculations following the rules of a-c theory, while the original pulse $C(t)$ does not. This is the reason why it is advantageous to introduce, instead of $C(t)$, the frequency spectrum $B(\omega)$, which is its perfect substitute. $B(\omega)$ is variously called the *periodogram* of the symmetrical signal $C(t)$, or its *Fourier spectrum*, or its *weight function* with regard to $\cos \omega t$.

5. Nominal Signal Duration and Nominal Cutoff.—After the spectrum of a certain pulse, for example the rectangular pulse $C(t)$, Sec. 3, has been obtained, it is not necessary to make new calculations in order to find the spectrum of another pulse of the same general shape but having different scales of time and voltage. Suppose, for example, that the second pulse $C'(t)$ spreads out in time four times as much as $C(t)$ and that its voltage amplitudes are three times as much. The spectrum $B'(\omega)$, of the second pulse, will be four times as narrow as the first one $B(\omega)$, since frequencies vary reciprocally with time; and the ordinates B' , being in volt-seconds, will be $3 \cdot 4 = 12$ times as great. The shape of $B'(\omega)$ otherwise will be the same as the shape of $B(\omega)$.

In dealing with a specific pulse, it is convenient to adopt as a unit of time a time duration physically related to that pulse, which the second is not. This is the same idea as introducing the time constant CR or L/R in the study of transients. Similarly, it is convenient to adopt a unit of frequency physically related to the spectrum, just as it is convenient to use ω_r as a unit in the study of a resonant (LRC) circuit.

A suitable unit of time¹ for the analysis of a symmetrical pulse is the *nominal signal duration* $2t_n$, a duration such that the area of a rectangle built on the central ordinate of the pulse and on the base $2t_n$ is equal to the algebraic value of the area between the pulse and the axis of time, Fig. 5.1. In a rectangular pulse, $2t_n$ is identical with the actual duration of the pulse $2t_1$, Fig. 3.1.

Similarly, a suitable unit of frequency is the *nominal cutoff* f_n , a frequency such that the area of a rectangle built on the initial ordinate ($\omega = 0$) of the spectrum and on the base $\omega_n = 2\pi f_n$ is equal to the algebraic value of the area between the spectrum and the axis $O\omega$, Fig. 5.2.

¹ H. A. WHEELER, The Interpretation of Amplitude and Phase Distortion in Terms of Paired Echoes, *Proc. I.R.E.*, **27**, 359, 1939.

From these definitions and the formulas (4.3) and (4.5) it can be shown that the following relation holds:

$$2f_n 2t_n = 1 \quad (5.1)$$

This relation is perfectly general and applies whatever the shape and size of the pulse. In the case of a rectangular pulse, (5.1) shows that the nominal cutoff is identical with the frequency of the auxiliary square wave of Fig. 3.3.

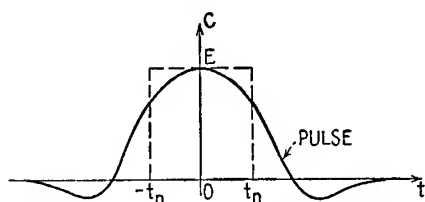


FIG. 5.1.—Nominal duration $2t_n$ of a symmetrical pulse $C(t)$.

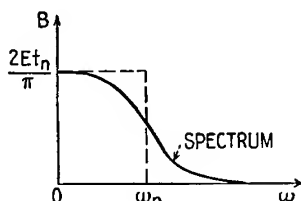


FIG. 5.2.—Nominal cutoff ω_n of a symmetrical pulse, derived from its spectrum $B(\omega)$.

6. Spectra of Various Symmetrical Pulses.—A list of some typical symmetrical signals, or pulses, $C(t)$ follows, with their spectra $B(\omega)$ as defined by formula (4.5). The amplitude of every pulse, for $t = 0$, is E volts; the amplitude of every spectrum for $\omega = 0$ is $B_0 = 2Et_n/\pi$ volt-sec; the nominal duration of every pulse is t_n ; the nominal cutoff of every spectrum is $f_n = 1/4t_n$. Taking t_n as unit of time, and f_n as unit of frequency, the following “absolute,” or “dimensionless,” variables x and y result,

$$x = \frac{t}{t_n} \quad y = \frac{f}{f_n} = \frac{\omega}{\omega_n} = 4t_n f$$

and these have been taken as abscissas in the diagrams of the pulse and of the spectrum, respectively.

Signal	Spectrum
1. Rectangular pulse Height E Duration $2t_n$	$B_0 \left[\frac{\sin (\pi y / 2)}{\pi y / 2} \right]$ (6.1)
2. Triangular pulse Peak value E Duration $4t_n$	$B_0 \left[\frac{\sin (\pi y / 2)}{\pi y / 2} \right]^2$ (6.2)
3. Half-cosine pulse Curve $E \cos x$ Duration πt_n	$B_0 \frac{1}{1 - (\pi y / 2)^2} \cos \left(\frac{\pi^2 y}{4} \right)$ (6.3)

4. Exponential pulse

$$\begin{array}{ll} \text{Curve } Ee^{-|x|} & B_0 \frac{1}{1 + (\pi y/2)^2} \\ \text{Duration } \infty & \end{array} \quad (6.4)$$

5. Probability pulse

$$\begin{array}{ll} \text{Curve } Ee^{-\pi x^2/4} & B_0 e^{-\pi y^2/4} \\ \text{Duration } \infty & \end{array} \quad (6.5)$$

These signals and spectra are given in Figs. 6.1 to 6.5.

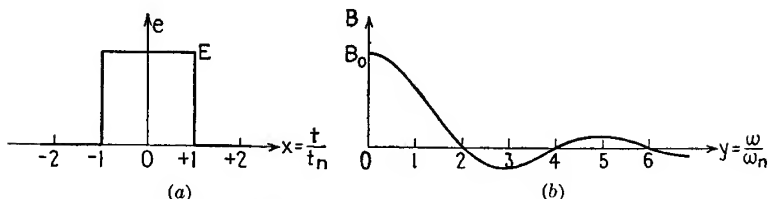


FIG. 6.1.—Rectangular pulse (a) and its spectrum (b); see (6.1).

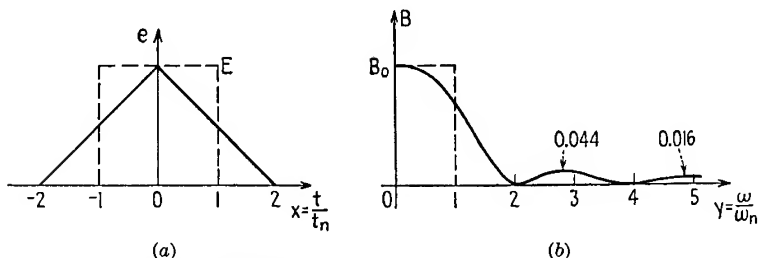


FIG. 6.2.—Triangular pulse (a) and its spectrum (b); see (6.2).

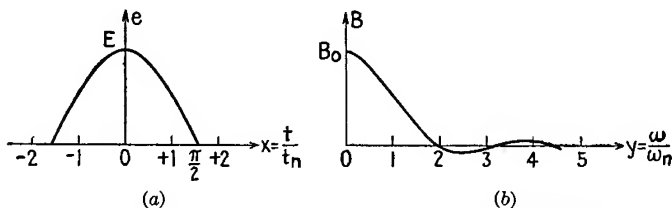


FIG. 6.3.—Half-cosine pulse (a) and its spectrum (b); see (6.3).

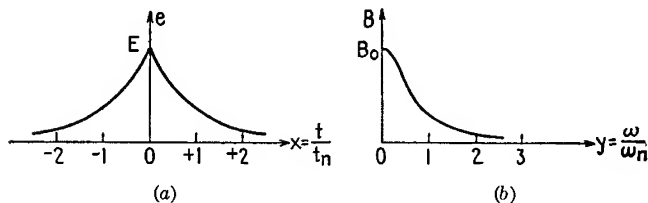


FIG. 6.4.—Exponential pulse (a) and its spectrum (b); see (6.4).

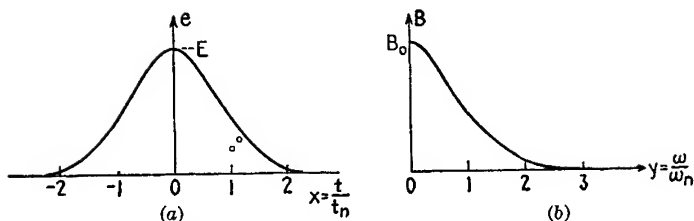


FIG. 6.5.—Probability pulse (a) and its spectrum (b); see (6.5).

7. Frequency Band for Transmission of Signal, Energy Spectrum.—From Figs. 6.1 to 6.5 all frequencies from zero to infinity are necessary for the exact reproduction of any one of the signals 1 to 5, Sec. 6.

Assume that the transducer is a low-pass filter of cutoff f_c and that it is required to determine the minimum band width (0 to f_c) which will transmit the signal with acceptable distortion.

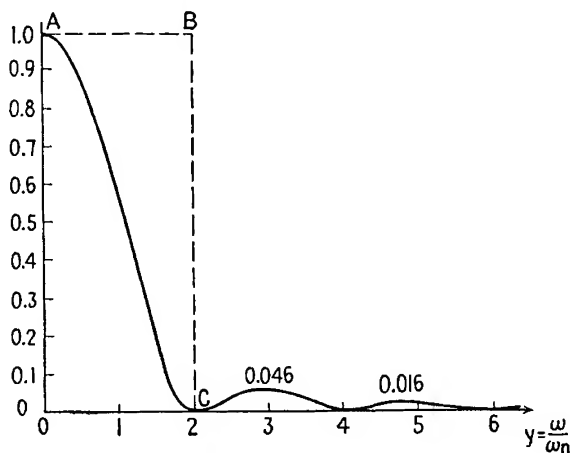


FIG. 7.1.—Energy spectrum for rectangular pulse.

No unique numerical definition can be given of what constitutes acceptable distortion. This will depend upon the type of receiver, whether intended for audio reception, visual reception, recording, printing, etc.

In many cases the distortion will be acceptable when the amplitude of all frequencies higher than cutoff f_c will be smaller than a prescribed fraction, say 1 per cent, of the maximum amplitude.

Another way of appreciating the relative importance of the component frequencies is to use the *energy spectrum*. This is the

plot of $[B(f)]^2$ as a function of frequency. For example, Fig. 7.1 represents the energy spectrum for a rectangular pulse. The whole area between that curve and the frequency axis represents the energy carried by the signal; in this specific case it is equal to the area $OABC$. The area under the curve limited by any two ordinates f_1, f_2 represents the energy carried by the frequencies lying in the f_1 - f_2 band. It can often be assumed that cutting off all frequencies above f_c will cause negligible distortion if the energy carried by these frequencies is a negligible fraction of the total energy carried by the signal.

Figures 6.1b to 6.5b show that the amplitude spectrum of the rectangular pulse decreases very slowly as the frequency increases; spectra 2, 3, 4 decrease faster; spectrum 5 decreases fastest of all. If a band of frequency 0 to $3f_n$ ($y = 3$) were transmitted uniformly, the probability pulse (No. 5) would suffer no appreciable distortion. To ensure the same quality of reproduction, a band about twice as wide would be required for pulses 2, 3, 4, and a band perhaps twenty times as wide for the rectangular pulse. This last condition is never present, and therefore sharp rectangular pulses are always distorted in transmission. It is usually considered sufficient to transmit uniformly the frequencies lying between $0.1f_n$ and $10f_n$.

8. Character of Pulse and Width of Its Spectrum.—In the Fourier series corresponding to a periodic symmetrical signal, discontinuities in the signal or in its derivatives involve a less rapid or more rapid decrease of B_n , Chap. IX. Similarly, it can be shown that if a nonperiodic signal is discontinuous (rectangular pulse of Fig. 6.1) the corresponding amplitude spectrum decreases for the high frequencies, of the order of $1/f$. If a signal is continuous but its first derivative is discontinuous (signals of Figs. 6.2, 6.3, 6.4) the amplitude spectrum decreases about as fast as $1/f^2$. If a signal and all its derivatives are continuous, the amplitude spectrum decreases faster than any power of $1/f$, that is, at least as fast as e^{-f} (signal of Fig. 6.4b and its spectrum, Fig. 6.4a). The spectrum of the probability pulse decreases as e^{-f^2} , which is still faster; and the spectra of signals such as those of Figs. 6.1b, 6.2b, 6.3b are identically zero for all frequencies higher than a certain limit, as shown in Figs. 6.1a, 6.2a, 6.3a.

9. Effect of Decreasing Duration of Pulse.—Consider the rectangular pulse of Fig. 6.1 and its spectrum. Transform Fig. 6.1 by multiplying the ordinates by m and dividing the abscissas by m as in Fig. 9.1; the two pulses of Fig. 9.1 have the same area.

The comparison of such pulses is often useful. (If they are current pulses, they transport the same electric charge; if they are force pulses plotted against displacement, they transport the same energy, etc.) The spectrum of the second pulse, Fig. 9.2, has the same amplitude $2Et_n$ as the first pulse and a nominal cutoff $m/4t_n$, m times greater. The height of the successive arches, which from formula (6.1) is approximately $2B_0/\pi y$, decreases as $1/f$ for the periodogram of the first pulse and as m/f for the second pulse. The

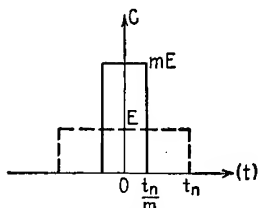


FIG. 9.1.—High and narrow rectangular pulse.

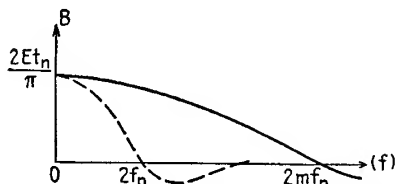


FIG. 9.2.—Spectrum of pulse of Fig. 9.1.

high frequencies are therefore m times more important in the high and narrow pulse than in the low and broad one (relative to the very low frequencies, of which the amplitudes $2Et_n$ are the same for both). Consequently, the band width required to pass either signal with a given amount of distortion is m times greater for the high and narrow pulse than for the low and broad one.

10. Smoothing a Signal.—This last result is a particular case of a more general relationship. When it is desired to “smooth out”

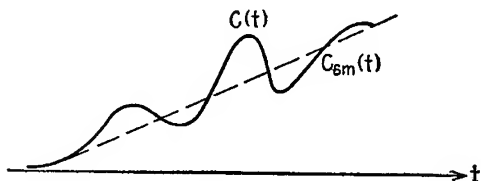


FIG. 10.1.—Relationship between $C(t)$ and $C_{sm}(t)$.

statistical data, a usual procedure is to replace the value of an item on any day by its average during, for example, 7 days, including 3 days before and 3 days after the given day. This is called a “sliding average.” Apply this procedure to a symmetrical signal $C(t)$, Fig. 10.1, by replacing $C(t)$ at every instant by its average value between the times $t - a$ and $t + a$, resulting in $C_{sm}(t)$, Fig. 10.1. It can be shown that the spectrum of $C_{sm}(t)$ is equal to the product of $B(\omega)$, the spectrum of $C(t)$, by $(\sin \omega a)/\omega a$.

A first consequence is that the new spectrum exhibits more marked oscillations than the original one, since $B(\omega)$ is multiplied by an oscillating factor that is periodically zero. While the signal is smoothed out, the spectrum becomes "wrinkled."

A second consequence is that the high frequencies are less important in the smoothed signal than in the original one; for if the amplitude spectrum of $C(t)$ diminishes as $1/\omega^k$, that of $C_{sm}(t)$ will diminish as $1/\omega^{k+1}$. Therefore, if a given amount of distortion is tolerated, a narrower frequency band will be needed to transmit the smoothed-out signal.

At first sight it might seem advantageous to choose a signal $e(t)$ demanding for its transmission a narrower band of frequencies. This overlooks the fact that the change of amplitude of $e(t)$ and its location in time are the essential factors constituting the information to be transmitted. If the smoothing process is repeated again and again, the information obtained becomes more and more vague and finally disappears. For maximum exactness of information an infinitely narrow pulse is ideal; this requires the use of all frequencies from zero to infinity, with equal weight given to all (as will be shown in Sec. 11). For maximum economy of frequencies an infinitely narrow band at f_n is desirable, and this corresponds to a harmonic voltage of constant amplitude for all times, which cannot transmit intelligence, Sec. 1. A compromise between these opposite requirements must be effected. One solution is the probability pulse, for which the signal and the periodogram both show a rapid decrease of amplitude. In each the area under the curve for time greater than $3t_n$ or frequency higher than $3f_n$ is only 1/5,000 of the total area. In practice, a signal approximating this shape will also approximate these qualities. For example, a signal having the shape of Fig. 6.4b has as its spectrum the curve of Fig. 6.4a; the amplitude of the frequency components decreases exponentially, *i.e.*, very rapidly.

11. Unit Function and Unit Pulse.—Consider a circuit containing a battery of emf E volts, a passive network Z , and a switch S , Fig. 11.1. If the circuit is closed at time $t = 0$, the potential difference at the terminals 1-2 of the network will be zero for all negative values of time $t < 0$ and E volts for $t > 0$, with a jump at $t = 0$, Fig. 11.2. When $E = 1$ volt, this function of t is called the *unit function* $U(t)$ [sometimes $1(t)$]; the function in Fig. 11.2 is thus $EU(t)$. This unit function plays a fundamental part in the work of Oliver Heaviside (about 1893) and of J. R. Carson.

Consider the two rectangular pulses described in Sec. 9. If m increases to infinity, the pulse becomes infinitely short and of infinitely great amplitude, the area under the pulse being the finite quantity $2Et_n$. When $2Et_n = 1$, the function of t (an infinitely sharp pulse of unit area) is called the *unit pulse* $P_1(t)$.

This function has been introduced in theoretical physics by P.

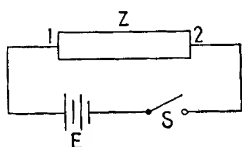


FIG. 11.1.—Sudden application of emf to network.

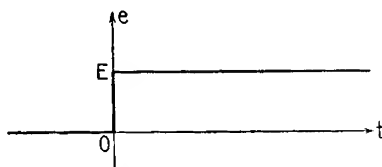


FIG. 11.2.—Voltage E times the unit function $U(t)$, yielding the function $EU(t)$.

A. M. Dirac (1926) and in circuit theory by G. A. Campbell (1927). It is being used more and more in the analysis of transients.

Consider a voltage equal to zero for $t < -t_n/m$ and to 1 volt for $t > t_n/m$ and represented for intermediate values of t by a straight line, Fig. 11.3. The derivative of such a function of time is equal to zero for $t < -t_n/m$ and $t > t_n/m$ and to $m/2t_n$ in between; it is a rectangular pulse, and its area is unity. If m increases to infinity, the *unit pulse* $P_1(t)$ may be considered as the *derivative* of the unit function $U(t)$ with regard to time.

If m increases to infinity, the spectrum, Fig. 9.2, approaches a horizontal line of ordinate $(1/\pi)2Et_n$. The unit pulse $P_1(t)$ has then for its spectrum the constant $1/\pi$, and all the component frequencies have the same weight.

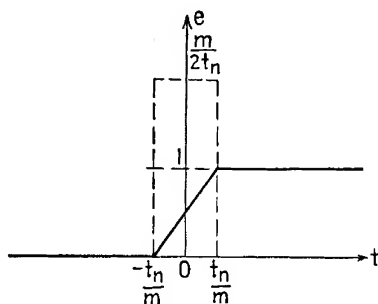


FIG. 11.3.—Voltage whose first derivative is a rectangular pulse.

If t_n continuously decreases toward zero and consequently f_n increases without limit and if E continuously increases so as to keep $2Et_n$ constant, each pulse in Figs. 6.1a to 6.5a becomes an instantaneous pulse of area $2Et_n$ and each spectrum in Figs. 6.1b to 6.5b becomes a horizontal line of height $(1/\pi)2Et_n$. Hence, an instantaneous pulse contains all frequencies with equal weight, a result often used.

12. Analysis of Antisymmetrical Pulse.—Instead of a symmetrical, or even, function of time, consider an antisymmetrical, or *odd*, function $S(t)$, such that $S(-t) = -S(t)$, Chap. IX, Sec. 3. Reasoning similar to that of Sec. 3 leads to the analysis of $S(t)$ into an infinity of components $\sin \omega t$ of *all* frequencies, each having an infinitely small amplitude $A(\omega) d\omega$, and the following formulas result instead of (4.2) and (4.5):

$$S(t) = \int_0^\infty A(\omega) \sin \omega t d\omega \quad (12.1)$$

$$A(\omega) = \frac{2}{\pi} \int_0^\infty S(t) \sin \omega t dt \quad (12.2)$$

$A(\omega)$ is called the *periodogram* of the antisymmetrical signal $S(t)$, or its *Fourier spectrum*, or its *weight function* with regard to $\sin \omega t$.

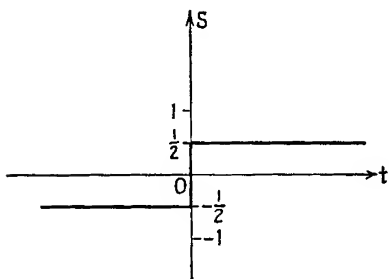


FIG. 12.1.—Unit step function $S_u(t)$.

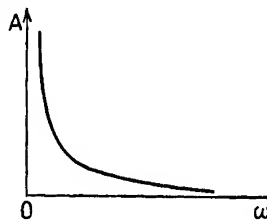


FIG. 12.2.—Spectrum $A(\omega) = 1/(\pi\omega)$ of unit step function.

An important antisymmetrical function is the *unit step function* $S_u(t)$, equal to $-\frac{1}{2}$ for $t < 0$ and to $+\frac{1}{2}$ for $t > 0$, Fig. 12.1. The spectrum of this function can be derived from that of the rectangular pulse and is

$$A(\omega) = \frac{1}{\pi\omega}$$

The amplitude of the harmonic component $\sin \omega t$ is $d\omega/\pi\omega$, inversely proportional to frequency, and the spectrum is a rectangular hyperbola, Fig. 12.2.

The definitions of t_n and f_n in Sec. 5 would be meaningless for an antisymmetrical signal. Arbitrary quantities t_0 and $f_0 = 1/4t_0$ may be used as units of time and frequency. When a signal of the general type $e(t) = C(t) + S(t)$ is analyzed, t_n and f_n for the symmetrical component $C(t)$ may be used for $S(t)$, also.

13. Spectra of Various Antisymmetrical Pulses.—A list of some typical antisymmetrical signals, or pulses, $S(t)$ follows with their spectra $A(\omega)$ as defined by formula (12.2). An arbitrary unit of time t_0 having been chosen, the following notations have been used, by analogy with the symmetrical case, Sec. 6:

$$f_0 = \frac{1}{4t_0}, \quad A_0 = \frac{2Et_0}{\pi}$$

$$x = \frac{t}{t_0} \quad y = \frac{f}{f_0} = \frac{\omega}{\omega_0} = 4t_0f$$

The absolute variables x and y have been taken as abscissas.

Signal	Spectrum
1. Step function $ES_u(t)$	
Jump at $t = 0$ from $-E/2$ to $E/2$	$A_0 \frac{1}{\pi y}$ (13.1)
2. Rectangular	
Amplitude $E/2$ Duration $2t_0$	$\frac{A_0}{2} \frac{1 - \cos \frac{\pi y}{2}}{\pi y/2}$ (13.2)
3. Exponential	
$(E/2)e^{-x}$ for $x > 0$ $-(E/2)e^x$ for $x < 0$	$\frac{A_0}{2} \frac{\pi y/2}{1 + (\pi y/2)^2}$ (13.3)
4. Sinusoidal (one period)	
Amplitude $E/2$ Duration $2\pi t_0$	$\frac{A_0}{2} \frac{\sin(\pi^2 y/2)}{1 - (\pi y/2)^2}$ (13.4)

The signals and spectra of (13.1) to (13.4) are given in Figs. 13.1 to 13.4.

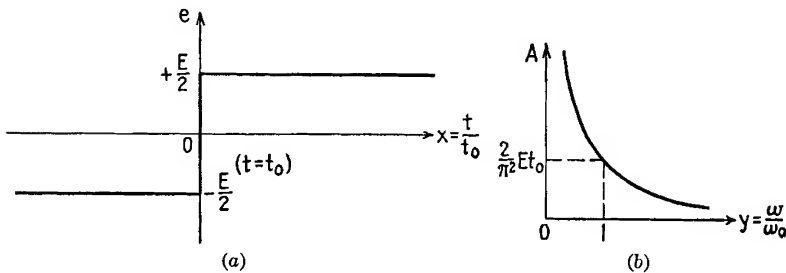


FIG. 13.1.—Step function $ES_u(t)$ (a) and its spectrum (b); see (13.1).

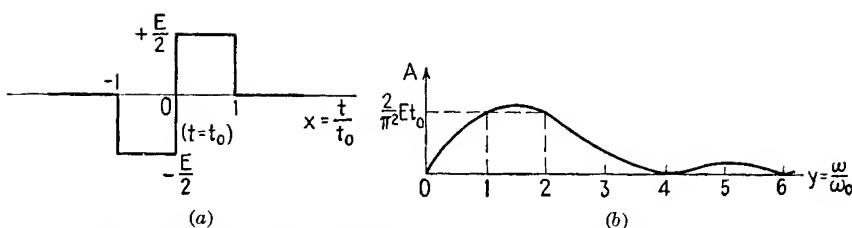


FIG. 13.2.—Antisymmetrical rectangular pulse (a) and its spectrum (b); see (13.2).

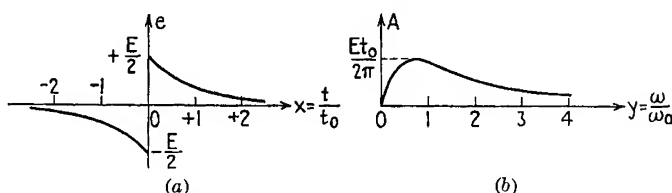


FIG. 13.3.—Antisymmetrical exponential pulse (a) and its spectrum (b); see (13.3).

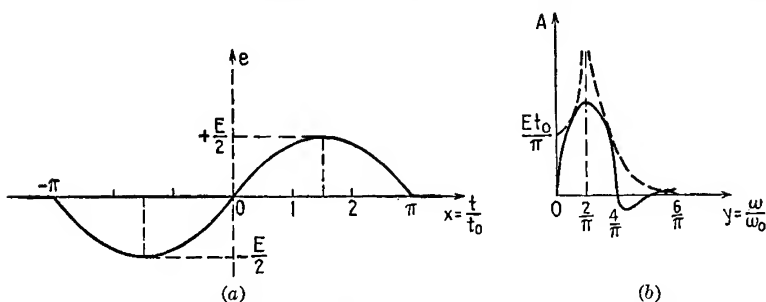


FIG. 13.4.—Antisymmetrical sinusoidal pulse (a) and its spectrum (b); see (13.4).

14. Signal Consisting of a Finite Number of Sinusoidal Cycles.—

The continuous passage from a nonperiodic function to a periodic one (inverse of the process considered in Sec. 3) is illustrated in the following problem:

Consider an antisymmetrical pulse composed of only one cycle of the function $(E/2) \sin x$ and equal to zero for $t < -\pi t_0$ and $t > \pi t_0$, Fig. 13.4. From formula (13.4) its spectrum is

$$A(\omega) = \frac{E t_0}{\pi} \frac{\sin \pi^2 y / 2}{1 - (\pi y / 2)^2} \quad (14.1)$$

Now consider a signal composed of $(2n + 1)$ cycles of the same function $(E/2) \sin x$. Such a signal, Fig. 14.1a, might be produced by an alternator of frequency $(1/2\pi t_0)$ operating for $(2n + 1)T$ sec, provided that there were no transients. The signals that have been considered thus far are called *d-c pulses*, in distinction to pulses

such as the present one, which are called *a-c pulses*. An a-c pulse is in general a carrier wave modulated by a d-c pulse, which here is a rectangular pulse.

From (12.2) the spectrum of the rectangular a-c pulse of duration $(2n + 1)2\pi t_0$ is

$$A_{2n+1}(\omega) = \frac{Et_0/\pi}{1 - (\pi y/2)^2} \sin \left[(2n + 1) \frac{\pi^2 y}{2} \right] \quad (14.2)$$

The spectrum $A_1(\omega)$ of a single sine-wave pulse is itself an amplitude-modulated wave, ω playing the role of time. The "period" of the carrier is $1/\pi t_0$, and the envelope is the curve

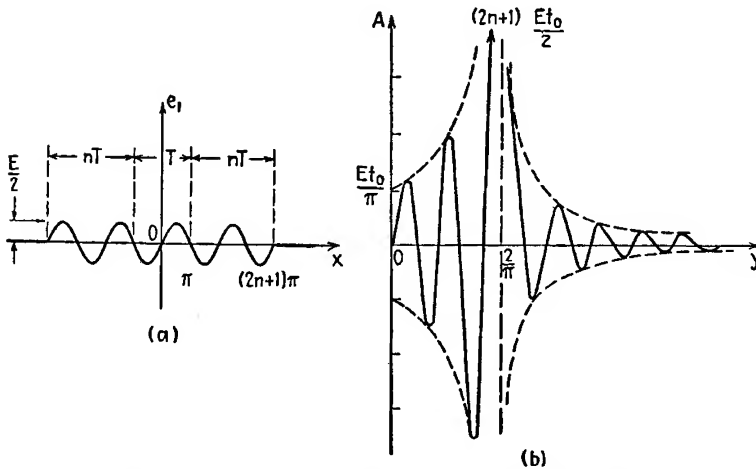


FIG. 14.1.—Antisymmetrical pulse of $(2n + 1)$ periods (a) and its spectrum (b).

$Et_0/\pi[1 - (\pi y/2)^2]$ together with its reflection in the $O\omega$ axis. The spectrum $A_{2n+1}(\omega)$ is composed of oscillations of $(2n + 1)$ times as short a "period," tangent to the same envelope. For $y = 2/\pi$, (14.1) and (14.2) become $\frac{0}{0}$; and, by the usual method for evaluating $\frac{0}{0}$, $A_1 = Et_0/2$ and $A_{2n+1} = (2n + 1)Et_0/2$. The spectra $A_1(\omega)$ and $A_{2n+1}(\omega)$ are plotted in Figs. 13.4b and 14.1b. The component frequency $f = 1/T$, corresponding to $y = 2/\pi$, is more and more preponderant in the spectrum as the number of a-c cycles in the signal increases. The carrier frequency does not stand out recognizably unless the signal contains at least 6 and preferably 10 or 12 periods. If the duration of the signal is decreased, the carrier frequency must be taken proportionately greater in order to maintain the same number of oscillations within the signal. This fact

has several important applications. A flash of light of 0.01 sec duration would contain about $5 \cdot 10^{12}$ oscillations. The corresponding spectrum would be practically an "instantaneous pulse," Sec. 11, with f instead of time as variable, centered at $f \doteq 5(10)^{14}$ cps. A signal that is an instantaneous pulse occurs at an instant that is perfectly well-defined, but its spectrum extends equally over all frequencies; on the other hand, a sinusoidal voltage $E \sin(t/t_0)$ has a perfectly well defined frequency but does not define any particular instant of time. A signal of *short* duration consisting of *many* oscillations of the carrier frequency is a good practical compromise between the two extreme theoretical cases.

Complementary properties such as those just mentioned are of great significance in contemporary physics.

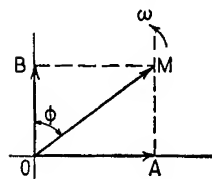
15. General Case.—In the general case, a signal $e(t)$ is the sum of a symmetrical and an antisymmetrical component,

$$e(t) = C(t) + S(t) \quad (15.1)$$

From (4.3) and (12.1),

$$e(t) = \int_0^\infty [B(\omega) \cos \omega t + A(\omega) \sin \omega t] d\omega \quad (15.2)$$

FIG. 15.1.—Rotating vectors corresponding to the harmonic components $B \cos \omega t$ and $A \sin \omega t$.



($B \cos \omega t + A \sin \omega t$) may be written as a cosine wave of amplitude M and phase angle $(-\phi)$, Fig. 15.1, with

$$M = \sqrt{A^2 + B^2} \quad (15.3)$$

$$\frac{B}{M} = \cos \phi \quad \frac{A}{M} = \sin \phi \quad \frac{A}{B} = \tan \phi \quad (15.4)$$

and, instead of (15.2),

$$e(t) = \int_0^\infty M \cos(\omega t - \phi) d\omega \quad (15.5)$$

Equation (15.5) means that $e(t)$ can be reconstructed by adding an infinity of waves of every frequency f , each having an infinitely small amplitude $M d\omega = \sqrt{A^2 + B^2} d\omega$ and a phase lag ϕ with regard to $\cos 2\pi ft$. The complex quantity $M d\omega e^{-j\phi}$ is called the *complex amplitude* of the component of angular velocity ω . $B(\omega)$ is called the *cosine spectrum* of the given signal $e(t)$, and $A(\omega)$ its *sine spectrum*. The complete Fourier analysis of $e(t)$ thus is summarized by *two* curves, representing two functions of ω , either $B(\omega)$ and $A(\omega)$, or $M(\omega)$ and $\phi(\omega)$.

By a generalization of the energy spectrum of a symmetrical pulse, defined in Sec. 7, the curve of $[M(f)]^2$ as a function of frequency is called the *energy spectrum* of the nonsymmetrical pulse $e(t)$.

16. Morse Dot, or Pulse.—As an example, a Morse dot of amplitude $2E$ and of duration t_n , beginning at $t = 0$, Fig. 16.1, is the sum of one symmetrical and one antisymmetrical rectangular

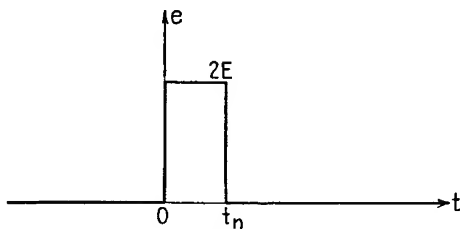


FIG. 16.1.—Unsymmetrical Morse dot.

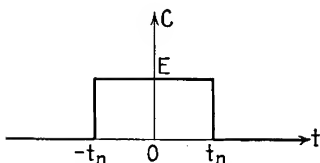


FIG. 16.2.—Symmetrical component of Morse dot of Fig. 16.1.

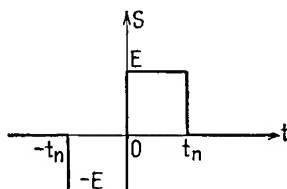


FIG. 16.3.—Antisymmetrical component of Morse dot of Fig. 16.1.

pulse, each of amplitude E , Figs. 16.2 and 16.3. From (6.1) and (13.2),

$$M(\omega) = \sqrt{A^2 + B^2} = \frac{2Et_n}{\pi y/2} \sqrt{\left[\sin \left(\frac{\pi y}{2} \right) \right]^2 + \left[1 - \cos \left(\frac{\pi y}{2} \right) \right]^2} \quad (16.1)$$

$$= \frac{4Et_n}{\pi y} \sqrt{2 \left[1 - \cos \left(\frac{\pi y}{2} \right) \right]}$$

and, from (15.4),

$$\tan \phi = \frac{1 - \cos (\pi y/2)}{\sin (\pi y/2)} = \tan \left(\frac{\pi y}{4} \right) \quad \phi = \frac{\pi y}{4} = \frac{t_n \omega}{2} \quad (16.2)$$

so that the relative amplitude M and phase ϕ of the $\cos \omega t$ component are as given by Figs. 16.4 and 16.5.

The Morse dot just analyzed could have been obtained by delaying a symmetrical Morse dot, of amplitude $2E$ and duration t_n , by $t_n/2$ sec. The comparison of Fig. 6.1b with Fig. 16.4 shows

that the amplitude spectra $M(\omega)$ of both pulses are the same but that a time delay $t_n/2$ sec corresponds to a phase lag $t_n\omega/2$ radians, proportional to frequency. This remark will be generalized in Sec. 22.

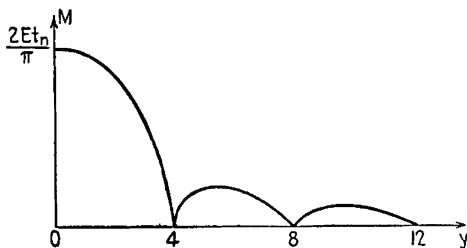


FIG. 16.4.—Amplitude diagram of Morse dot of Fig. 16.1.

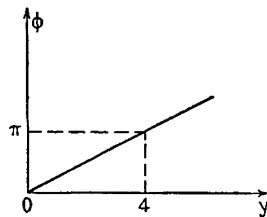


FIG. 16.5.—Phase diagram of Morse dot of Fig. 16.1.

17. Fourier Transform.—It is convenient to represent a harmonic wave $E \cos \omega t$ by a rotating vector as in a-c theory or by the corresponding complex exponential $Ee^{j\omega t}$, of which $E \cos \omega t$ is the real part. The wave $M \cos (\omega t - \phi)$ will then be represented by $Me^{j(\omega t - \phi)}$.

The amplitude $M(\omega)$ defined in Sec. 15 can be expressed as

$$M(\omega) = M_0 e^{-a(\omega)} \quad (17.1)$$

where M_0 (in volt-seconds) is $1/\pi$ times the area $2Et_n$ of the pulse or of its symmetrical part. The analysis of a pulse $e(t)$, (15.5), can then be written as

$$e(t) = \int_0^\infty M_0 e^{-a-j\phi} e^{j\omega t} d\omega \quad (17.2)$$

The complex function $M(\omega)/\phi(\omega)$, or $M_0 e^{-a-j\phi}$, is called the (complex) *Fourier transform* $E(\omega)$ of the pulse $e(t)$. It is its weight function with regard to $e^{j\omega t}$. The quantities a (nepers) and ϕ (radians) are both functions of frequency.

Conversely, the Fourier transform is related to the pulse by

$$\begin{aligned} E(\omega) &= B(\omega) - jA(\omega) \\ &= \frac{2}{\pi} \int_0^\infty e(t) e^{-j\omega t} dt \end{aligned} \quad (17.3)$$

The knowledge of the two frequency characteristics $a(\omega)$, $\phi(\omega)$ is equivalent to the knowledge of the signal $e(t)$, and vice versa. Thus the *one* function $e(t)$ corresponds to the *two* functions $a(\omega)$, $\phi(\omega)$ [or to the *two* functions $B(\omega)$, $A(\omega)$, Sec. 15]. This is possible

only because the two functions (a, ϕ) [or the two functions (B, A)] are not independent, as will be shown in Sec. 21.

It is shown in Chap. IX, Sec. 12, that the properties of a network in the harmonic steady state can also be expressed by two frequency characteristics $\alpha(\omega)$ and $\beta(\omega)$ (attenuation α and phase lag β). It is thus possible to describe these two widely different physical entities, the input signal and the network, in comparable terms. This makes it possible similarly to analyze their product, which is the output signal.

DISTORTION OF SIGNAL GOING THROUGH A TRANSDUCER

18. Fourier Analysis of Output Signal.—As stated in Sec. 1, the problem is to find the shape of the signal $e_2(t)$ that reaches a given receiver, when a given nonperiodic signal $e_1(t)$ is applied at the input terminals of a given transducer, Fig. 1.4. Let

$$\frac{E_2}{E_1} = A(\omega) = |A|/\underline{-\beta} = \epsilon^{-\alpha-j\beta} \quad (18.1)$$

be the complex voltage ratio of the transducer when connected to the receiver. If the input voltage is harmonic of frequency f , $e_1(t) = E_1 \cos \omega t$, the output voltage is, Chap. IX, Sec. 12,

$$e_2(t) = E_1 \epsilon^{-\alpha} \cos(\omega t - \beta) \quad \text{or} \quad E_1 \epsilon^{-\alpha-j\beta} e^{j\omega t} \quad (18.2)$$

If the input signal $e_1(t)$ is nonperiodic, it is shown in Sec. 15 that it is equivalent to the sum of an infinity of harmonic components of all frequencies, each of them infinitely small, and that the component of frequency f is

$$M_0 \epsilon^{-a_1} \cos(\omega t - \phi_1) d\omega \quad \text{or} \quad M_0 \epsilon^{-a_1-j\phi_1} e^{j\omega t} d\omega \quad (18.3)$$

where $M_0 = 2Et_n/\pi$. The harmonic output voltage corresponding to this input component of $e_1(t)$ will be

$$M_0 \epsilon^{-a_1-j\phi_1} \epsilon^{-\alpha-j\beta} e^{j\omega t} d\omega \quad (18.4)$$

Putting

$$a_1 + \alpha = a_2 \quad \phi_1 + \beta = \phi_2 \quad (18.5)$$

(18.4) becomes

$$M_0 \epsilon^{-a_1-\alpha} \epsilon^{-j\phi_1-j\beta} e^{j\omega t} d\omega = M_0 \epsilon^{-a_2-j\phi_2} e^{j\omega t} d\omega \quad (18.6)$$

Thus, from (18.5), the a_2 of the output signal is the sum of the a_1 of the input signal and the attenuation α of the transducer; the ϕ_2 of the output signal is the sum of the ϕ_1 of the input signal and the phase lag β of the transducer.

From (17.2), the Fourier transform of the input signal $e_1(t)$ is $E_1(\omega) = M_0 e^{-\alpha_1 - j\phi_1}$, and from (18.6) the Fourier transform of the output signal $e_2(t)$ is $E_2(\omega) = M_0 e^{-\alpha_1 - j\phi_1}$, whereas from (18.1) the voltage ratio of the transducer (connected to the receiver) is $A(f) = e^{-\alpha - j\beta}$. Equation (18.6) can therefore be written more simply as

$$E_2(\omega) = A(\omega) E_1(\omega) \quad (18.7)$$

Thus, the Fourier transform of the output signal is the product of the voltage ratio of the transducer and the Fourier transform of the input signal.

This rule is essentially the same as that, (18.1), concerning the harmonic input and output voltages of the transducer,

$$E_2 = A(\omega) E_1 \quad (18.8)$$

which is not surprising, since the Fourier transform of a non-periodic signal is nothing more than the complex amplitude of each of its components, with the differential factor $d\omega$ deleted, Sec. 17.

19. Derivation of Output Signal from Input Signal.—To obtain the output signal the following three operations must be performed, parallel to those described in Chap. IX for an input periodic voltage:

1. Find the Fourier transform $E_1(\omega)$ of the input signal (Fourier analysis).
2. Multiply $E_1(\omega)$ by the voltage ratio $A(\omega) = e^{-\alpha - j\beta}$ of the transducer; this gives the Fourier transform $E_2(\omega)$ of the output signal, 18.7.
3. Obtain the output signal $e_2(t)$ from the knowledge of its Fourier transform (Fourier synthesis).

This process is strikingly similar to performing a numerical calculation with a table of logarithms: (1) Find the logarithms of the data. (2) Perform on them some simple arithmetical operation. (3) Find the number of which the last number obtained is the logarithm.

Steps 1 and 3 require the evaluation of the definite integrals (17.3) and (17.2). This involves elaborate mathematical methods. The results of many such calculations have been tabulated in "Fourier Integrals for Practical Applications," by G. A. Campbell and R. M. Foster (A.T.T.Co.).

20. Response of Network to Unit Pulse.—The unit instantaneous pulse $P_1(t)$ has for spectrum the constant $1/\pi$, Sec. 11. Putting $E_1(\omega) = 1/\pi$ (volt-seconds), (18.7) becomes

$$E_2(\omega) = \frac{1}{\pi} A(\omega) \quad (20.1)$$

This relation brings out a remarkable connection between the steady-state properties and the transient properties of a given network. The complex voltage ratio $E_2/E_1 = A(\omega)$, obtained from ordinary a-c theory, describes the behavior of the network in the harmonic steady state for any given frequency. On the other hand, the application of a unit instantaneous pulse to the network results in an output signal $P_2(t)$, usually called a transient. Equation (20.1) signifies that $(1/\pi)A(\omega)$ is the Fourier transform of the transient response $P_2(t)$.

For example, consider the transducer and receiver of Fig. 20.1. The voltage ratio is

$$\frac{E_2}{E_1} = \frac{1/j\omega C}{R + j\left(\omega L - \frac{1}{\omega C}\right)} = A(\omega) \quad (20.2)$$

The signal $P_2(t)$ of which $1/\pi A(\omega)$ is the Fourier transform is obtained by finding the value of the integral (17.2), which here is

$$P_2(t) = \int_0^\infty \frac{1}{\pi} \frac{e^{j\omega t} d\omega}{(1 - \omega^2 LC) + j\omega CR} \quad (20.3)$$

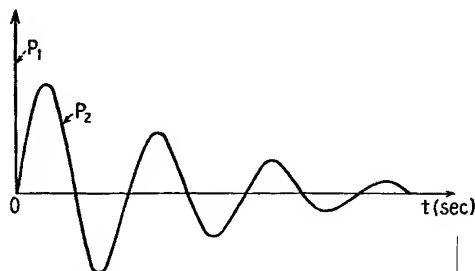


FIG. 20.2.—Response of circuit of Fig. 20.1 to unit pulse $P_1(t)$, or shock excitation.

Formula (448.1) of Foster's tables, written in the present notation, gives

$$P_2(t) = \frac{1}{\pi} \frac{1}{\omega_1 LC} = e^{-\frac{R}{2L}t} \sin \omega_1 t \quad (\text{volts}) \quad (20.4)$$

with

$$\omega_1 = \sqrt{\frac{1}{LC} - \left(\frac{R}{2L}\right)^2} \quad (20.5)$$

The signal $P_2(t)$, for t positive, is the damped sinusoid given by (20.4) and shown in Fig. 20.2. The fact that one extremely short knock thus sets a resonant circuit in oscillation is the well-known phenomenon of *shock excitation* on which early radio communication by damped waves was based.

21. Interdependence of Attenuation and Phase Characteristics of Network.—The response of a network to a unit pulse may help in

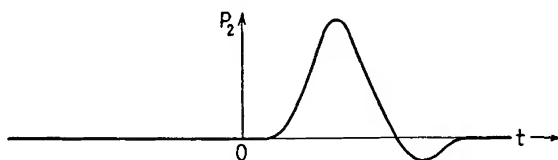


FIG. 21.1.—Example of an output voltage pulse, $P_2(t)$.

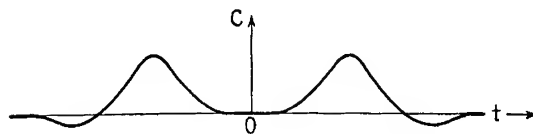


FIG. 21.2.— C component of $P_2(t)$.

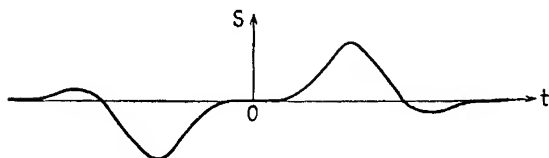


FIG. 21.3.— S component of $P_2(t)$.

understanding the fact mentioned in Sec. 17, that the two frequency characteristics of a pulse, $a(\omega)$, $\phi(\omega)$, or the two frequency characteristics of a network, $\alpha(\omega)$, $\beta(\omega)$, are not independent.

Suppose that a transducer, originally at rest (no current in any branch), receives at $t = 0$ an instantaneous input voltage pulse $P_1(t)$. The output voltage $P_2(t)$ may begin to change at $t = 0$ or later if the transducer contains a filter or a transmission line, which introduces a delay, but surely not earlier than $t = 0$, Fig. 21.1.

If $P_2(t)$ is analyzed into a C and an S component, the right-hand parts of C and S will be identical, Figs. 21.2 and 21.3, whereas the left-hand parts will be the image of each other in the time axis.

If C were known, we could immediately draw the S curve, and vice versa.

Now $1/\pi$ times the voltage ratio $A(\omega)$, or $1/\pi[A_r(\omega) + jA_i(\omega)]$, is the Fourier transform of $P_2(t)$. The spectrum of the C component of $P_2(t)$ is $A_r(\omega)/\pi$, and the spectrum of the S component of $P_2(t)$ is $A_i(\omega)/\pi$. Since the C component and the S component of $P_2(t)$ are not independent, their spectra $A_r(\omega)$ and $A_i(\omega)$ cannot be independent.

Since $A_r(\omega)$ and $A_i(\omega)$, real part and imaginary part of $A(\omega)$, are not independent, the amplitude and the phase angle of $A(\omega)$ cannot be independent. From (17.1), $A(\omega) = e^{-\alpha-j\beta}$, where the function $\alpha(\omega)$ is the attenuation of the transducer and the function $\beta(\omega)$ is its phase angle. These functions therefore are not independent, hence cannot both be chosen at random; if, in designing a transducer, it is desired that its attenuation have a given variation as function of frequency, its phase angle is automatically determined, and vice versa.¹

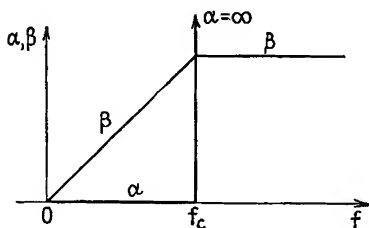


FIG. 21.4.—Characteristics of ideal low-pass filter.

Suppose, for example, it were desired to design a low-pass filter, Fig. 21.4, such that, in the transmission band, α is zero and β proportional to frequency; and that, in the attenuation band, α is

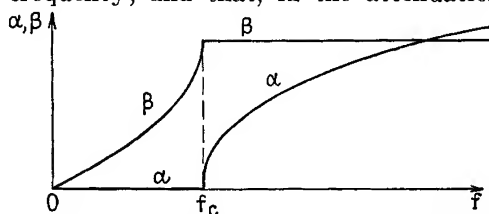


FIG. 21.5.—Actual characteristics of low-pass filter.

infinite and β a constant. These conditions are not compatible. If they are assumed, they result in a response $P_2(t)$ to a unit pulse that is not zero for $t < 0$; that is, an output voltage would exist before any input voltage has been applied. Actually, for the low-pass filter of Chap. VIII, Sec. 5, α is zero below cutoff, but then β is not proportional to frequency; above cutoff, β is a constant, but then α is not infinite, Fig. 21.5.

22. Conditions for No Distortion.—The Fourier analysis provides at once the conditions that should be fulfilled by an ideal

¹ These general statements are somewhat qualified in a more advanced theory.

transducer, in order that it may transmit an applied voltage of any shape without distortion. Consider one sinusoidal component of $e_1(t)$, of angular velocity ω ; it will be attenuated α nepers and be subject to a phase lag of β radians, or to a time delay $t_d = (\beta/\omega)$ sec, Sec. 16.

If α and t_d are constants independent of frequency, all components of $e_1(t)$ will be attenuated in the same ratio and delayed the same amount of time. Hence $e_2(t)$ will be a perfect copy of $e_1(t)$, delayed t_d sec.

There are, therefore, two conditions for no distortion:

1. The attenuation α introduced by the network must be independent of frequency.

2. The delay t_d introduced by the network must be independent of frequency.

This second condition may be expressed also by saying that the phase angle β introduced by the network must be proportional to frequency: $\beta = 2\pi t_d f$.

Moreover, multiplying $A(\omega)$ by a factor $e^{-j\pi} = -1$ merely reverses the output voltage, so that any multiple of π can be added to the phase lag β without introducing distortion.

23. Different Types of Distortion.—The networks here considered (transducer and receiver, Fig. 1.4) may contain resistances, inductances, either self or mutual, and capacitances, as well as transmission lines and vacuum tubes. It has been assumed that the complex voltage ratio $A(\omega)$ could be obtained by applying Kirchhoff's laws to the network in the harmonic steady state, such relations being used as

$$e_R = Ri \quad e_L = j\omega Li \quad r_p i_p = e_p + \mu e_g \quad \cdots \quad (23.1)$$

A network is said to be *linear* when all coefficients such as R , L , C , M , r_p , μ , . . . , can be considered as independent of the amplitudes of the harmonic voltages and currents involved. It results from the theory of linear differential equations with constant coefficients that, when a harmonic voltage or current of frequency f is applied to a *linear* network, all the voltages and currents in the network are also harmonic and of the same frequency f .

A coil with an iron core and a vacuum tube so biased that the curved portions of the characteristics come into play are examples of *nonlinear* systems. If a harmonic voltage of frequency f is applied to such a system, the input current is not harmonic; and if a nonperiodic signal is applied, the output signal is distorted in a

very complicated way, as is seen, for example, in the study of detection, Chap. XXI. Distortion caused by nonlinear networks is called *nonlinear distortion*, and the present theory does not apply to it.

Assuming a linear network, no distortion of the input signal will occur if the network characteristics satisfy the conditions

$$\left. \begin{array}{l} \alpha = \text{const.} \\ t_d = \text{const.} \end{array} \right\} \text{or} \left. \begin{array}{l} \alpha = \text{const.} \\ \beta = 2\pi t_d f + n\pi \end{array} \right\} \quad (23.2)$$

If the actual characteristics of the linear transducer (connected to a definite linear receiver) differ from the ideal ones, Figs. 23.1

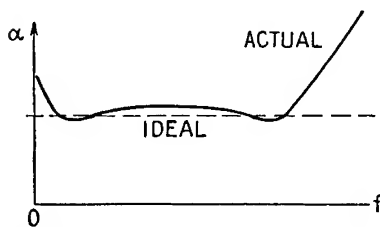


FIG. 23.1.—Actual and ideal attenuation characteristics.

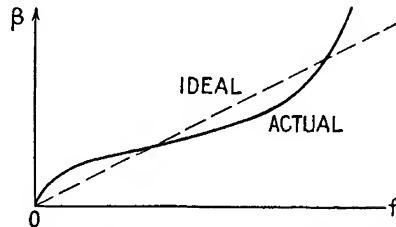


FIG. 23.2.—Actual and ideal phase characteristics.

and 23.2, distortion will occur. The American Standards Association¹ recommends the following definitions:

Frequency distortion is that form of distortion in which the change is in the relative magnitudes of the different frequency components of a wave, provided that the change is not caused by nonlinear distortion.

Delay distortion is that form of distortion which occurs when the phase angle of the transfer impedance (or of the voltage ratio) with respect to two chosen pairs of terminals is not linear with frequency within a desired range, thus making the time of transmission or delay vary with frequency in that range.

Since the magnitudes of the different frequency components are controlled by the attenuation $\alpha(f)$ and their time delays by the phase $\beta(f)$, both types of distortion just defined depend upon the shape of some frequency characteristic, either that of Fig. 23.1 or that of Fig. 23.2, so that the choice of the term "frequency distortion" for one of them seems unfortunate. Nonconformist writers such as Guillemin or Wheeler use the terms "amplitude distortion" and

¹ American Standard Definitions of Electrical Terms, ASA C42-1941, American Institute of Electrical Engineers.

"phase distortion" and have logic on their side. In what follows we shall use the official definitions.

In the reproduction of voice and music, delay distortion, if not extreme, is not recognized by the ear; therefore the designer aims at obtaining a flat attenuation characteristic and lets the phase angle take care of itself. Conditions are entirely different in television, where even very slight delay distortion causes objectionable distortion of the picture. New types of filters and amplifiers have been designed to keep the time delay constant throughout the very broad frequency band used in the technique of the cathode-ray oscillograph.

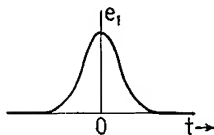


FIG. 24.1.—Input signal.

24. Effect of Linear Distortion on Symmetrical Pulse.—Let the phase angle $\beta(\omega)$ of the voltage ratio be sometimes greater and sometimes less than an angle $t_d\omega$ proportional to frequency. Let

$$\beta = t_d\omega + \beta_1 \quad (24.1)$$

Then, from (18.1),

$$A(\omega)\epsilon^{-\alpha-j\beta} = \epsilon^{-j\omega t_d}\epsilon^{-\alpha-j\beta_1} \quad (24.2)$$

The first factor $\epsilon^{-j\omega t_d}$ delays the input signal by the amount t_d , Sec. 22. The second factor, where α and β_1 are functions of fre-

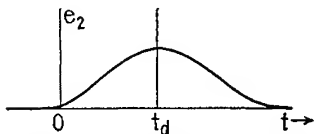


FIG. 24.2.—Signal delayed by t_d and flattened by frequency distortion.

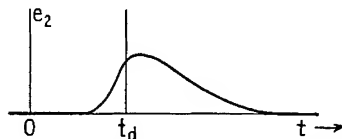


FIG. 24.3.—Total effect of linear distortion; skewing is due to delay distortion.

quency, has a real part and an imaginary part:

$$\epsilon^{-\alpha-j\beta_1} = (\epsilon^{-\alpha} \cos \beta_1) - j(\epsilon^{-\alpha} \sin \beta_1) \quad (24.3)$$

The real part is the spectrum of a symmetrical pulse; the coefficient of $(-j)$ is the spectrum of an antisymmetrical pulse. A symmetrical signal, passing through a linear transducer, is thus affected in general in three ways:

1. It is delayed by t_d sec.
2. As a result of α not being a constant (frequency distortion), the shape of the signal is altered, but its symmetry about the vertical $t = t_0$ is preserved.

3. As a result of β not being equal to ωt_d (delay distortion), an unsymmetrical component is added, or the signal is skewed.

In general, (2) broadens and flattens the pulse symmetrically, and (3) makes the upstroke shorter (steeper) and the downstroke longer (flatter). The total effect is called linear distortion. A typical case is shown in Figs. 24.1 to 24.3.

25. Method for Obtaining Shape of Output Signal When Distortion Is Slight.—The harmonic-analysis method enables the output voltage corresponding to a given input voltage to be found, provided that certain mathematical integrations can be effected, as explained in Sec. 19 and exemplified in Sec. 20. Several important problems in telegraphy and television have been treated in this way.

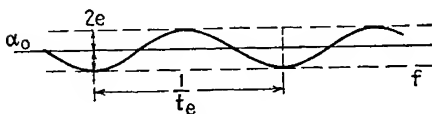


FIG. 25.1.— α -Characteristic of system I.

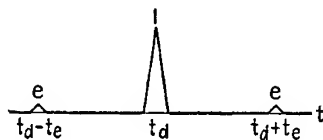


FIG. 25.2.—Signal at the output of system I.

In practical problems, the variations of the input voltage with time and the α - and β -characteristics of the network are usually obtained by automatic recording or by plotting measurements and therefore are represented by graphs, for which a rigorous analytical expression is not available. An ingenious method for obtaining the output signal, the *method of paired echoes*, has been devised by H. A. Wheeler. The method consists essentially in replacing the analytical expression by a Fourier development that is its equivalent over the useful range of frequencies.

Wheeler begins by investigating two fictitious systems. System I has frequency distortion only; system II has delay distortion only.

The α -characteristic of system I is assumed to be a sinusoid of amplitude $2e$ nepers ($e < 1$) of which one cycle occupies a frequency band $1/t_e$ cps, Fig. 25.1. The β -characteristic, being ideal, is a straight line of slope $1/2\pi t_d$. If the attenuation α were constantly zero, a very short pulse of amplitude 1, Fig. 25.2, would be reproduced exactly after a delay of t_d sec. Wheeler shows that owing to the frequency distortion present this main output signal will be accompanied by two echoes, one arriving t_e sec earlier than the main signal and the other t_e sec later, both having the amplitude e , Fig. 25.2.

The β -characteristic of system II is assumed to be a sinusoid of amplitude $2e'$ nepers ($e' \ll 1$), of which one cycle occupies $1/t_e'$ cps, oscillating about a straight line of slope $1/2\pi t_d$, Fig. 25.3. The attenuation α is constantly zero. Owing to the delay distortion present, the main signal of amplitude 1, arriving after t_d sec, is accompanied by two echoes, one arriving t_e' sec earlier than the

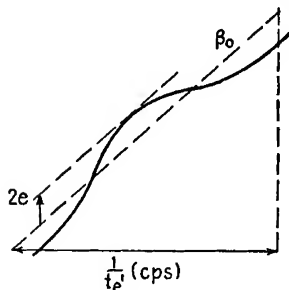


FIG. 25.3.— β -Characteristic of system II.

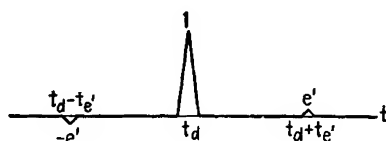


FIG. 25.4.—Signal at the output of system II.

main signal and the other t_e' sec later, one having the amplitude e' and the other the amplitude $(-e')$, Fig. 25.4.

Systems I and II are fictitious in that their α - and β -characteristics are not consistent. Characteristics of an actual system will always be consistent.

For example let the α -characteristic of an actual system be as in Fig. 25.5. There will always be an upper limit f_c to the frequencies transmitted by the system. The shape of the character-

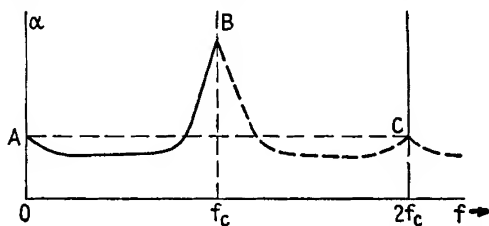


FIG. 25.5.—Actual α -characteristic.

istics above f_c is therefore immaterial. It may be assumed that in the frequency band f_c to $2f_c$ the α -characteristic is the image BC of the characteristic AB . The periodic function of frequency represented by $ABC \dots$ will have a Fourier development

$$\alpha_0 + \alpha_1 \cos \left(\frac{2\pi}{2f_c} \right) + \dots \quad (25.1)$$

It will usually be sufficient to consider only the two first terms. Let $\alpha_1 = 2e\alpha_0$. The output signal then will consist of a main signal (the input signal attenuated by α_0 nepers) accompanied by two echoes of amplitude e times as small, preceding and following the main signal by $1/2f_c$ sec. If this approximation is not sufficient, other Fourier components may be used, each one contributing a smaller and smaller pair of echoes.

In the same way the actual β -characteristic will be referred to a straight line $\beta = 2\pi t a f$ going through it, and the difference between the actual and the ideal characteristic will be developed in a Fourier series. Each successive Fourier component will contribute a smaller and smaller pair of echoes of opposite amplitudes, Fig. 25.4.

Supplementary Reading

Treatment of Fourier Integrals in Textbooks

- J. R. CARSON: "Electric Circuit Theory and the Operational Calculus," Chap. XI, McGraw-Hill Book Company, Inc., New York, 1926.
T. E. SHEA: "Transmission Networks and Wave Filters," Chap. XI, D. Van Nostrand Company, New York, 1929.
F. A. GUILLEMIN: "Communication Networks," Vol. II, Chap. XI, John Wiley & Sons, Inc., New York, 1935.
R. V. CHURCHILL: "Fourier Series and Boundary Value Problems," Chaps. V, VI, McGraw-Hill Book Company, Inc., New York, 1941.

Table of Integrals

- G. A. CAMPBELL and R. M. FOSTER: "Fourier Integrals for Practical Applications," American Telephone and Telegraph Company, 1931. (Lists 763 pairs of Fourier transforms.)

Method of Paired Echoes

- H. A. WHEELER: The Interpretation of Amplitude and Phase Distortion in Terms of Paired Echoes, *Proc. I.R.E.*, **27**, 359, 1939.

CHAPTER XIX

PRINCIPLES OF MODULATION

1. The Modulating Signal and Types of Modulation.—In its nontechnical sense, the word “modulate” means to conform to a certain proportion; to temper or regulate; to vary or inflect in tone. In its technical sense as used in radio, modulate means to change or alter the properties of a wave or current in some systematic manner in order to convey a message. The wave or current whose properties are altered or modulated is called the carrier wave or carrier current. The “wave” that causes and controls the alteration is called the “modulating wave.” The process of producing the modulated or systematically altered wave is called a “process of modulation” or, more simply, “modulation.” The word modulation is also used as a name for the changes produced in the process of modulation.

An electromagnetic wave is employed as a carrier in radio communication. The modulating wave in radiotelephony is a varying electric current obtained from a microphone. In television the modulating wave is a complex electrical signal produced in part by a tube that operates essentially as a photoelectric cell. In radiotelegraphy the modulating wave is a pattern of dots and dashes.

The fundamental importance of modulation is that it enables communication to be carried on at frequencies different from those of the modulating wave and by means of power derived from sources other than that of the modulating wave. The frequency of the carrier wave and the nature of the modulation process determine the frequencies present in the modulated wave and the relative power at the different frequencies.

For a given type of modulation, the properties of the modulated wave are the same regardless of the particular apparatus employed to produce the modulation. Hence it is fitting to discuss the *principles* of modulation without regard to the workings of any specific apparatus for producing modulation. It is the purpose of this chapter to study the nature of the modulation process, with major emphasis on the frequencies and the band widths involved.

In Chap. XX on Methods of Modulation the operation of present-day equipment for producing a modulated wave is described.

In most applications, the modulating wave is too irregular in form to be described by an explicit equation. However, the performance of a system may be judged or predicted on the basis of its response to sinusoidal modulating waves covering a band of frequencies. For example, spoken words are reproduced with sufficient intelligibility to be useful if the reproducing apparatus, from microphone to loudspeaker or earphones, has a fairly uniform response from about 300 to 3,000 cps. Phase distortion is not important, considerable nonlinear distortion may be tolerated, and frequency distortion is sometimes purposely employed to emphasize the higher frequencies in order to improve the intelligibility. For satisfactory reproduction of music, a wider frequency range is required and the permissible nonlinear distortion is considerably reduced. Ordinary radiobroadcasts cover an audio-frequency range of 50 to 5,000 cps or more. The wider the range, the more natural the reproduction, provided that the nonlinear distortion is kept low enough to avoid unpleasant aural effects. Satisfactory television requires a band several million cycles wide and a reproducing system having little frequency and phase distortion, depending upon the desired clearness of the picture. The required frequency range for telegraphy increases with the sending speed in words per minute, but for manual keying the necessary band width is only a few hundred cycles.

In practically all communication systems, the reproduction of the modulating wave is not perfect. There is always some frequency, phase, or nonlinear distortion somewhere in the system. Sometimes this distortion is unavoidable or is tolerated in order to meet practical specifications involving cost, weight, size, life, etc. In other cases, "distortion," *i.e.*, departure from the modulating wave, is deliberately introduced, for example to compensate for certain defects in a microphone or camera tube or to secure some aural or visual effect in the reproduction. The high or the low frequencies or both are sometimes emphasized in sound reproductions to secure a more natural or more distinct reproduction or to suit the fancy of the listener. Also, the volume range, *i.e.*, the range of intensity, may be compressed to accommodate the limitations of a reproducing device or locale, or to prevent the intensity from becoming so weak as to be overridden by noise, or to prevent excessively loud sounds, which may cause undesirable distortion or

annoyance. In television work the wave obtained from the camera is altered for many purposes, including that of changing the brightness and contrast of the reproduced image. In short, many "tricks" or refinements, involving alterations before or after the process of modulation, are employed for the purpose of securing some desired effect in the system as a whole. In the remainder of this chapter *the modulating wave including all changes made previous to the process of modulation, i.e., the wave fed into the modulator, will be called the "modulating signal."* The modulation process involves changing the carrier wave in synchronism with and in proportion to the changes in the modulating signal. The changes in the carrier wave produced by modulation are detected in the receiver and used to control the currents by means of which the final reproduction is made.

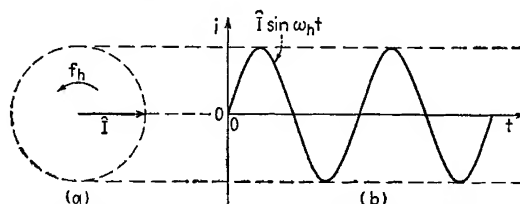


FIG. 1.1.—Graphical representation of an r-f current as a function of time by means of a rotating vector.

The radio frequency or carrier frequency is higher than the frequencies of the modulating signal and is designated by the subscript h . The frequencies of the modulating signal are lower than the carrier frequency and are designated by the subscript l . The carrier frequency may be higher than the modulating frequency by a factor of 3 or 4 in some carrier-current telephony applications and by a factor of thousands in short-wave radio.

When there is no modulation, the carrier-frequency current in the output of the equipment designed to produce modulation may be described by the expression

$$i = \hat{I} \sin \omega_h t \quad (1.1)$$

where \hat{I} represents the amplitude of the current and ω_h represents its angular frequency, which is 2π times f_h , the carrier frequency. It is assumed, for convenience, that the current is zero at zero time. The sine function is chosen arbitrarily; the cosine function could be used as well. As discussed in Chap. I, the instantaneous value of a sinusoid may be represented by the vertical projection of a

rotating vector, as in Fig. 1.1. In Fig. 1.1a, a vector representing the amplitude \hat{I} of the current of (1.1) is shown in its position at $t = 0$. Its vertical projection then is zero. The vector is imagined to rotate counterclockwise at an angular velocity $\omega_h = 2\pi f_h$ radians/sec. The effect of modulation is to vary such a vector in length, in angular motion, or in both length and angular motion. If only the length is changed, the wave is amplitude-modulated. If only the angular motion is changed, the wave is angle-modulated. These two types of modulation are basic types. Each type of modulation requires appropriate modulation and detection equipment.

2. Amplitude Modulation.—Amplitude modulation is defined here as the process (or the results of the process) of varying the

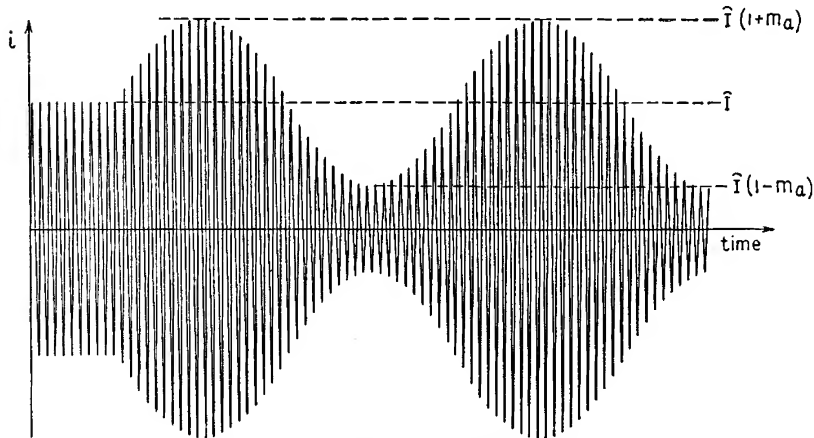


FIG. 2.1.—Amplitude-modulated current.

amplitude of the carrier in synchronism with and in proportion to the variations in the modulating signal. This definition describes the ideal of amplitude modulation. Most devices do not attain the ideal called for by this definition, but the process and results nevertheless are called “amplitude modulation.”

The modulating signal is seldom if ever a sinusoidal signal, but the operation of the system as a whole may be evaluated on the basis of its response to sinusoidal signals covering the necessary frequency band, as discussed in Sec. 1. Figure 2.1 shows the results of sinusoidal amplitude modulation, or amplitude modulation by a sinusoidal modulating signal. At first, the amplitude is constant; during this time, there is no modulating signal. Thereafter the

amplitude is shown to be varying sinusoidally about the value that it had before modulation. The sinusoidal variation in amplitude is proportional to the instantaneous value of a sinusoidal modulating signal. The time intervals between successive zero values of the modulated r-f current are equal and are unchanged by amplitude modulation.

The maximum fractional change in amplitude is known as the *degree of modulation* and is designated as m_a . As illustrated in Fig. 2.1 the amplitude of the modulated r-f current varies sinusoidally between the values $\hat{I}(1 + m_a)$ and $\hat{I}(1 - m_a)$ at the frequency of the modulating signal. The variation of the amplitude with time is expressed by the multiplier $(1 + m_a \sin \omega t)$, where

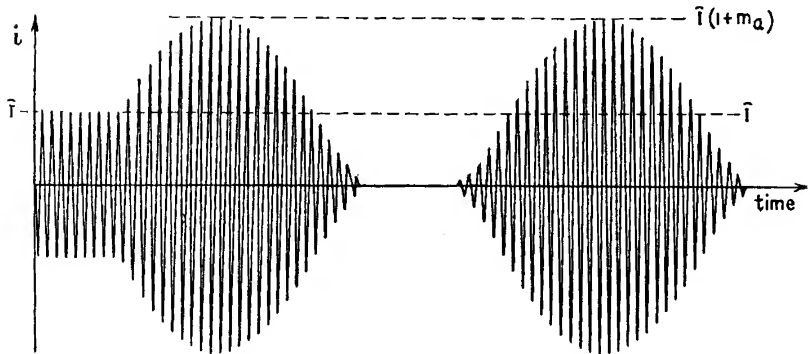


FIG. 2.2.—Overmodulated current.

$\omega_i = 2\pi f_i$, f_i being the frequency of the modulating signal. Then (1.1), altered to include the multiplier expressing the change in amplitude, becomes

$$i = \hat{I}(1 + m_a \sin \omega_i t) \sin \omega_c t \quad (2.1)$$

The degree of modulation m_a is proportional to the amplitude of the modulating signal. For normal modulation, m_a may have any value between unity and zero. When m_a has the value unity, the amplitude of the modulated wave varies between twice the no-signal amplitude and zero. This condition is known as 100 per cent modulation. The percentage of modulation for any degree of amplitude modulation is m_a times 100.

With most modulating devices, *overmodulation* with a sinusoidal modulating signal results in a waveform like that shown in Fig. 2.2. The amplitude is more than doubled during part of each modulating period, while the r-f oscillations disappear entirely during another

part. The variation in the amplitude of the r-f wave does not correspond exactly to the waveform of the modulating signal. This results in a distorted reproduction in the receiver.

3. Side Frequencies and Side Bands Due to Amplitude Modulation.—By use of the relation

$$\sin x \sin y = \frac{1}{2} \cos (x - y) - \frac{1}{2} \cos (x + y)$$

(2.1) may be converted into the form

$$i = \hat{I} \sin \omega_h t + \frac{m_a \hat{I}}{2} \cos (\omega_h - \omega_l) t - \frac{m_a \hat{I}}{2} \cos (\omega_h + \omega_l) t \quad (3.1)$$

The amplitude-modulated current expressed by (2.1) and illustrated in Fig. 2.1 thus is seen to be the sum of three components, each having a sinusoidal waveform. The *center* component on the frequency scale, or the *carrier* component, has the same frequency and amplitude as the current before modulation. The other two components are called the *side-frequency components*, the $\omega_h - \omega_l$ term representing the *lower side frequency* and the $\omega_h + \omega_l$ term representing the *upper side frequency*. Each of the side-frequency components has an amplitude equal to $m_a/2$ times the amplitude of the carrier and differs in frequency from the carrier by an amount equal to the frequency of the modulating signal.

The three components of a sinusoidally amplitude-modulated wave are shown as functions of time in Fig. 3.1, where it is assumed that $m_a = 1$ so that each side component has an amplitude half as great as the carrier component. In one period of the sinusoidal modulating signal, the upper side-frequency wave has one cycle more and the lower side-frequency wave has one cycle less than the carrier wave. Curve *d*, the modulated wave, is obtained when curves *a*, *b*, *c* are added together. Since $m_a = 1$, the amplitude of wave *d* varies between twice the amplitude of the unmodulated carrier and zero. If m_a had been assumed less than unity, the amplitudes of waves *b* and *c* would have been proportionately smaller and the envelope of *d* would not expand and contract to the extent shown.

An *amplitude spectrum* of the three component currents of Fig. 3.1 is shown in Fig. 3.2. Lines erected along a frequency axis at $f_h - f_l$, f_h , $f_h + f_l$ are proportional in length to the amplitudes indicated by (3.1) with m_a equal to unity. The lengths of the side-frequency spectrum lines would be proportionately less for smaller values of m_a .

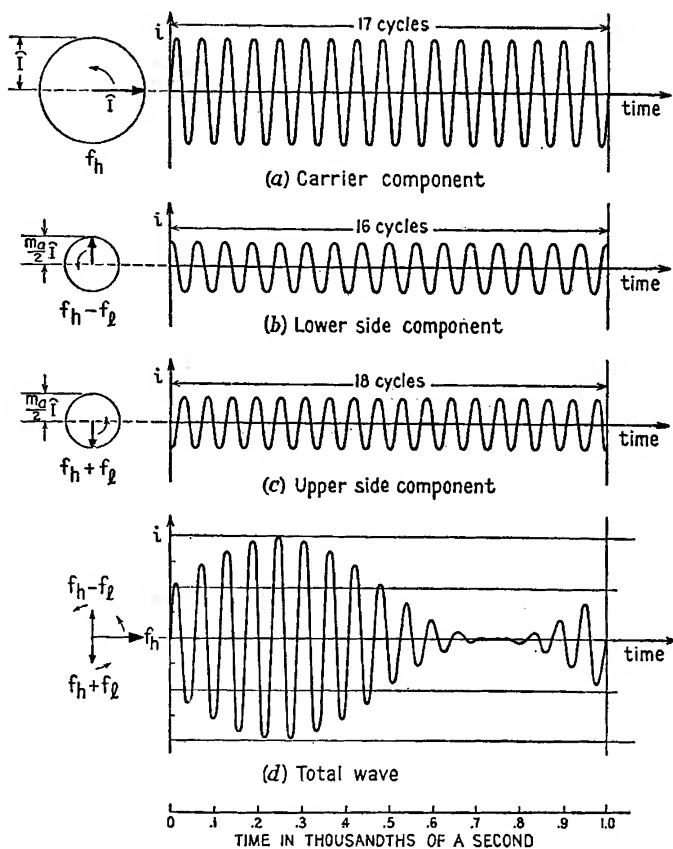


FIG. 3.1.—Diagrams representing a carrier of 17,000 cps sinusoidally amplitude modulated at a frequency of 1,000 cps.

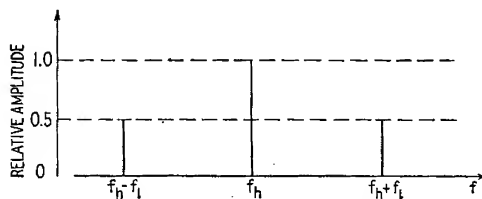


FIG. 3.2.—Amplitude spectrum of a sinusoidally modulated current with m_a equal to unity.

A complex modulating wave may be regarded as the sum of sinusoidal components of different frequencies. Each component causes two side-frequency waves of the type just described. Therefore, during modulation by a complex sound, many side frequencies lie on either side of the carrier in groups called *side bands*. Within each side band the individual side-frequency waves wax and wane in accordance with the variations in the modulating signal. The total band width in cycles/second is equal to twice the highest modulation frequency.

In the standard broadcast band, carrier frequencies are assigned 10 keps apart. Therefore the nominal channel width is 10 keps, and the nominal maximum modulating-signal frequency is 5,000 cps in this band. Broadcast transmitters, however, can and do modulate at frequencies higher than 5,000 cps.¹ Interference between stations due to overlapping side bands is not usually noticed because adjacent channels are not assigned to stations located in geographically adjacent districts.

The envelope of the overmodulated wave of Fig. 2.2 is non-sinusoidal. This envelope would be produced in an "ideal" amplitude-modulation process by a modulating signal having the same waveform as the envelope. Since a nonsinusoidal signal contains harmonics, Chap. IX, it follows that many pairs of side frequencies are contained among the components of this overmodulated wave. There is one pair of side frequencies for every harmonic component of the modulating signal. Hence a highly overmodulated wave has a very wide spectrum and may cause objectionable interference between stations.

4. Power during Amplitude Modulation.—The average power due to a current is proportional to the average of the square of the current. When the amplitude-modulated current illustrated in Fig. 2.1 is squared, the squared current varies with time as shown in Fig. 4.1. During modulation the *increases* in power exceed the *decreases* in power. Therefore the *average* power of the modulated wave is *greater* than the average power of the unmodulated wave. Since the carrier component of a modulated wave is unchanged by modulation (unless overmodulation occurs), the carrier power is unchanged and all the added power is in the side bands.

If the current of (3.1) is maintained in a resistance R , the average power due to the carrier component is $\hat{I}^2 R/2$, and that due

¹ R. F. GUY, Engineering Factors Involved in Relocating WEAF, *R.C.A. Rev.*, 5, 455, 1941.

to each of the side-frequency components is $m_a^2 \hat{I}^2 R / 8$. The total average power is the sum of the average power due to the currents of the three different frequencies and is

$$P = \left(\frac{\hat{I}^2 R}{2} + m_a^2 \frac{\hat{I}^2 R}{8} + m_a^2 \frac{\hat{I}^2 R}{8} \right) \\ = \frac{\hat{I}^2 R}{2} \left(1 + \frac{m_a^2}{2} \right)$$

where $\hat{I}^2 R / 2$ is the power before modulation. The power due to the carrier component being $P_c = \hat{I}^2 R / 2$,

$$P = P_c \left(1 + \frac{m_a^2}{2} \right) \quad (4.1)$$

During 100 per cent sinusoidal modulation the average power in the modulated wave is 150 per cent of the average carrier power, or

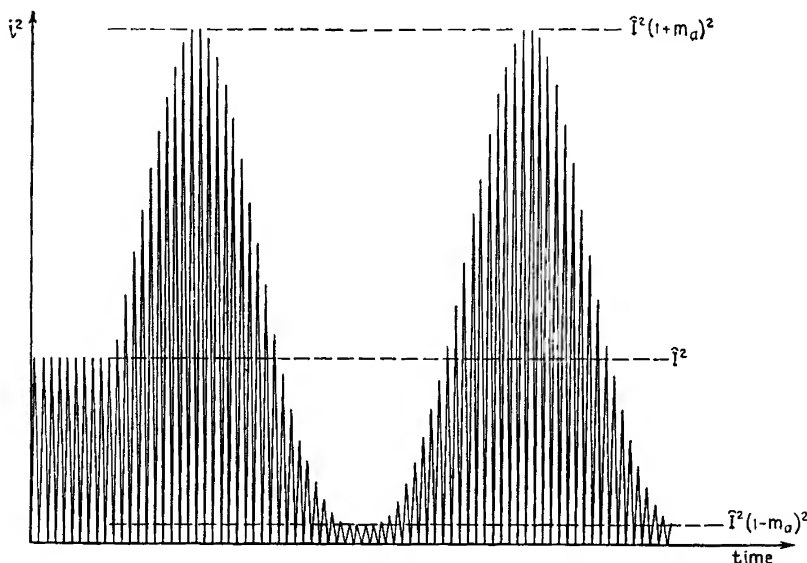


FIG. 4.1.—Square of sinusoidally amplitude-modulated current.

of the power in the absence of modulation. An rms ammeter under this condition would indicate a current equal to $\sqrt{1.5}$, or 1.225 times the current in the absence of modulation. However, sustained modulation with a sinusoidal signal is most unusual and probably would be encountered only in testing. In practice, the signal usually is complex and irregular, and consequently the total power

represented by the amplitude-modulated wave customarily fluctuates. The 50 per cent increase in power and 22.5 per cent rise in ammeter reading may be regarded as theoretical upper limits, reached in practice only occasionally if at all.

The intelligence may be said to lie in the side bands of an amplitude-modulated wave. The greater the degree of modulation, the greater the amount of energy placed in the sidebands. Figure 4.2 shows what might be considered as a possible energy spectrum averaged over a period of time. In general, there is more energy in the lower frequencies of an audio-frequency modulating wave than in the higher frequencies. The energy in the side frequencies of an amplitude-modulated wave is proportional to the energy in the modulating signal. Hence the side frequencies nearer the carrier frequency have on the average more energy than the side

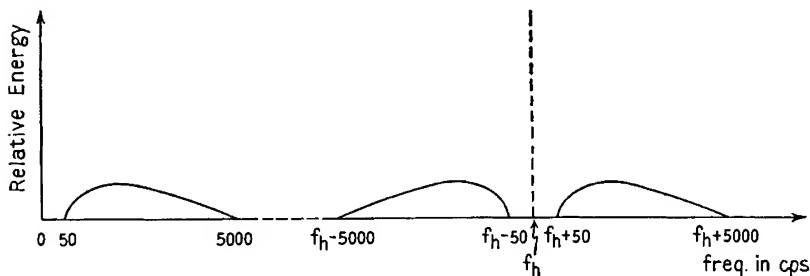


FIG. 4.2.—Possible time-average energy spectrum of a complex modulating wave and of the r-f wave produced by amplitude modulation.

frequencies farther from the carrier frequency. In amplitude-modulation practice, the carrier is modulated to as high a degree as possible without the occurrence of objectionable overmodulation, in order to maximize the power that represents the intelligence.

5. Representation of an Amplitude-modulated Wave by Rotating Vectors.—The components in (3.1) with m_a equal to unity are represented by the vectors shown at the left in a , b , c , Fig. 3.1. The initial positions and relative lengths of these vectors are obtained from (3.1) by setting t equal to zero. The speed of rotation of each vector in revolutions/second is equal to the frequency of the component wave that it represents.

Since the wave in d is obtained by adding waves a , b , c , it follows that, if the vectors representing a , b , c are first added, the successive vertical projections of the resultant will yield curve d .

To visualize the addition of the vectors, imagine them arranged like the hands of a clock, but with each vector turning counter-

clockwise at its respective speed. The observer may then imagine himself riding around with the carrier-frequency vector. In effect, this subtracts the frequency f_c from the speed of all three vectors. To the rotating observer, the carrier-frequency vector would appear motionless; the lower side-frequency vector would appear to rotate

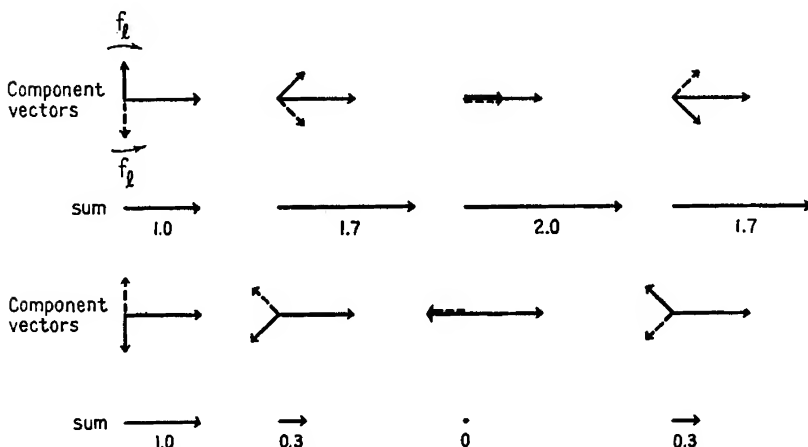


FIG. 5.1.—Component vectors and resultant vector at eight equal intervals during one modulation period for sinusoidal amplitude modulation, $m_a = 1$.

clockwise at f_l rps, while the upper side-frequency vector would appear to rotate counterclockwise at f_l rps. At $t = 0$, the observer would see the three vectors in the first position in Fig. 5.1. At intervals of one-eighth of a modulation period thereafter, the vectors

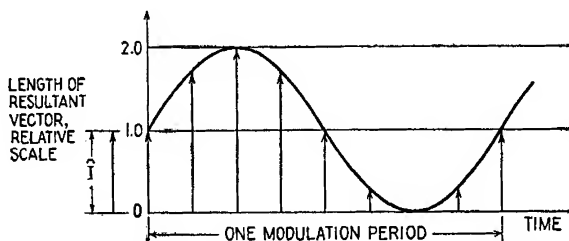


FIG. 5.2.—Length of resultant during one modulation period (sinusoidal modulation, $m_a = 1$).

would be seen in the other seven positions. The side-frequency vectors always make equal angles with the carrier-component vector. The resultant of the three vectors is always collinear with the carrier-component vector. The resultant is shown on the second and fourth lines of Fig. 5.1. The length of the resultant is plotted

as a function of time in Fig. 5.2. The length varies sinusoidally. During 100 per cent modulation, the length alternately doubles and shrinks to zero. The degree of lengthening and shrinking would be less for smaller degrees of modulation.

Since the resultant vector remains collinear with the carrier-component vector, it necessarily rotates at the frequency f_c rps, rotating uniformly as it lengthens and shrinks. Its vertical projection is zero at equal intervals of time. Therefore, as shown in Figs. 2.1 and 3.1*d*, the amplitude-modulated wave crosses the time axis at equal intervals. The resultant vector goes through one complete cycle of lengthening and shrinking in a time interval equal to the period of the modulating signal. Its angular motion remains uniform.

6. Angle Modulation.—In the discussion of amplitude modulation, the modulated current is represented by the use of a vector of varying length, rotating at constant speed. There is no difference in angle between the vector representing the amplitude-modulated wave and the vector representing only the carrier wave. However, there are other methods of modulation in which the representative vector is altered only in its angle with respect to the position it would have in the absence of modulation. These methods are designated by the general term *angular*, or *angle modulation*. The representative vector does not change in length but rotates at a speed alternately above and below the speed at which it would rotate in the absence of modulation. It is alternately ahead of or behind the position it would have in the absence of modulation. In practice, angle modulation is commonly called “frequency modulation.” However, systems that differ greatly in their operation are so commonly labeled frequency-modulation, or “FM,” systems that for purposes of discussion the term frequency modulation should be defined. Two types of angle modulation are defined and described in the following sections.

Phase modulation as defined here is a type of angle modulation in which the *phase shift* (the angle between the representative vector and the position it would occupy if modulation were absent) is proportional to the instantaneous value of the modulating signal. Figure 7.2*b* shows the result for the sinusoidal signal of Fig. 7.2*a*.

Frequency modulation as defined here is a type of angle modulation in which the *frequency shift* (the deviation of the instantaneous frequency¹ from that of the unmodulated wave) is proportional to

¹ As defined by (7.3), p. 630.

the instantaneous value of the modulating signal. Figure 7.2c shows the result for the sinusoidal signal of Fig. 7.2a.

Any change in angular motion from a uniform angular velocity involves *simultaneously* both *phase shift* and *frequency shift*, since advancing the position of one rotating vector with respect to another involves at least a temporary speed-up so that one vector may advance ahead of the other. This speed-up means an increase in frequency since frequency is proportional to angular velocity. Hence any shift in phase involves a shift in frequency, and vice versa.

In practice, many devices are a compromise and strictly speaking produce neither phase modulation nor frequency modulation as here defined. However, they do vary the angle or the angular motion of the representative vector and hence should be called "angular modulators" or "angle modulators." Actually, they are called "frequency modulators," and the term "frequency modulation," or "FM," is used loosely in practice to designate any form of angle modulation. In the practical application of angle modulation the apparatus employed usually produces a wave resembling that which would be produced by ideal frequency modulation as here defined.

Phase and frequency modulation are similar in that each involves a change in the angular motion of the representative vector. For a sinusoidal modulating signal of period $T = 1/f_i$, Fig. 7.2a, maximum frequency deviation and maximum phase deviation occur alternately $T/4$ sec apart, maximum phase deviation of the phase-modulated wave coinciding in time with maximum frequency deviation of the frequency-modulated wave.

In practice, the phase shift or frequency shift may not always be proportional to the amplitude of the modulating signal. Just as changes in the modulating wave may be tolerated or purposely made to secure certain effects, Sec. 1, so departures from the ideal in the process of angle modulation may be tolerated for reasons of economy or to secure some desired effect in the transmission or reproduction of the desired intelligence.

In angle modulation or in the special types of angle modulation here defined as phase and frequency modulation, the modulated wave may be represented by the equation

$$i = \hat{I} \sin \theta \quad (6.1)$$

where \hat{I} is a constant amplitude and θ is an angle that increases nonuniformly with time. The angle θ is the total angle swept

over up to the time t , measured from an arbitrary position at $t = 0$, by a nonuniformly rotating vector of constant length. Angle modulation leaves \hat{I} unchanged in magnitude but causes θ to increase at a nonuniform rate. Several devices to produce angle modulation are discussed in Chap. XX.

Much of the difficulty accompanying the initial study of phase and frequency modulation (or, more generally, angle modulation) will be overcome if three things are kept in mind. (1) The trigonometric sine in (6.1) involves an *angle*, not a speed. (2) The angle θ of (6.1) is not equal to ωt when ω is not a constant.¹ (3) Angular velocity ω and hence frequency f are proportional to the rate of change of the angular displacement θ .

7. Phase Modulation.—With no modulation, the representative vector rotates at constant speed and the angle θ of (6.1) increases

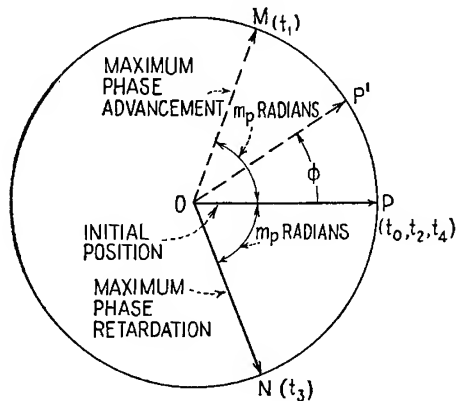


FIG. 7.1.—Motion of a vector representing a phase-modulated current resulting from sinusoidal modulation, with f_h rps subtracted from its motion.

uniformly at the rate ω_h radians/sec. During phase modulation, the ever-increasing angle θ is alternately advanced “ahead of schedule” and retarded “behind schedule” by amounts that are at every moment proportional to the successive positive or negative values of a modulating signal. If flashlight pictures of the representative vector were taken regularly f_h times per second, the vector in the absence of modulation would be seen always in a fixed position, such as OP in Fig. 7.1. During phase modulation, the snapshots would show the vector moving from one side to the other of OP , one such position being OP' . A time plot of the successive

¹ It will be recalled that distance is not equal to the product of instantaneous speed and time when a change of speed occurs.

values of the angle between the vector and the position OP would have the same waveform as that of the modulating signal.

This effect may be visualized also by imagining two vectors, one rotating uniformly, representing an unmodulated wave, the other rotating at the nonuniform speed, corresponding to phase modulation. The observer then may imagine himself riding around on the steadily rotating, or unmodulated, carrier-frequency vector, looking in the direction of rotation. The phase-modulated vector would then be alternately ahead of or behind the observer, in accordance with the positive or negative values of the modulating signal. The angle measuring the position of the vector representing the modulated wave, with respect to the vector representing the unmodulated wave, or with respect to the rotating observer, may be called ϕ . In phase modulation this angle is proportional to the instantaneous value of the modulating signal.

For a sinusoidal modulating signal, $A \sin \omega t$, Fig. 7.2a, the angle ϕ varies sinusoidally, so that the total angle θ measured from the starting point departs sinusoidally from the value given by $\omega_h t$. If the greatest departure, or the greatest value of ϕ , is m_p radians, the instantaneous departure is $m_p \sin \omega t$, and the total angle θ increases with time according to the relation

$$\theta = \omega_h t + m_p \sin \omega t \quad (7.1)$$

A series of snapshots taken regularly f_h times per second would show the vector executing simple harmonic angular motion at the frequency f_i between the two positions OM and ON , Fig. 7.1. These extreme positions are m_p radians removed from the position OP . The maximum phase shift m_p is directly proportional to the amplitude A of the modulating signal. Maximum advancement, at position OM , is reached at times t_1, t_5, \dots , Fig. 7.2a, when the modulating wave has its maximum positive value. Maximum retardation, at position ON , is reached at times t_3, t_7, \dots , when the modulating wave has its maximum negative value.

Substitution of (7.1) in (6.1) gives an equation for a phase-modulated current resulting from a sinusoidal modulating signal,

$$i = \hat{I} \sin (\omega_h t + m_p \sin \omega t) \quad (7.2)$$

as represented in Fig. 7.2b.

In practice, the maximum phase shift m_p may have a value ranging from a small fraction of a radian to many radians. In

amplitude modulation, "100 per cent modulation" is a limit to distortion-free modulation. There is no such limit in phase modulation. Distortion is absent when the angle ϕ is propor-

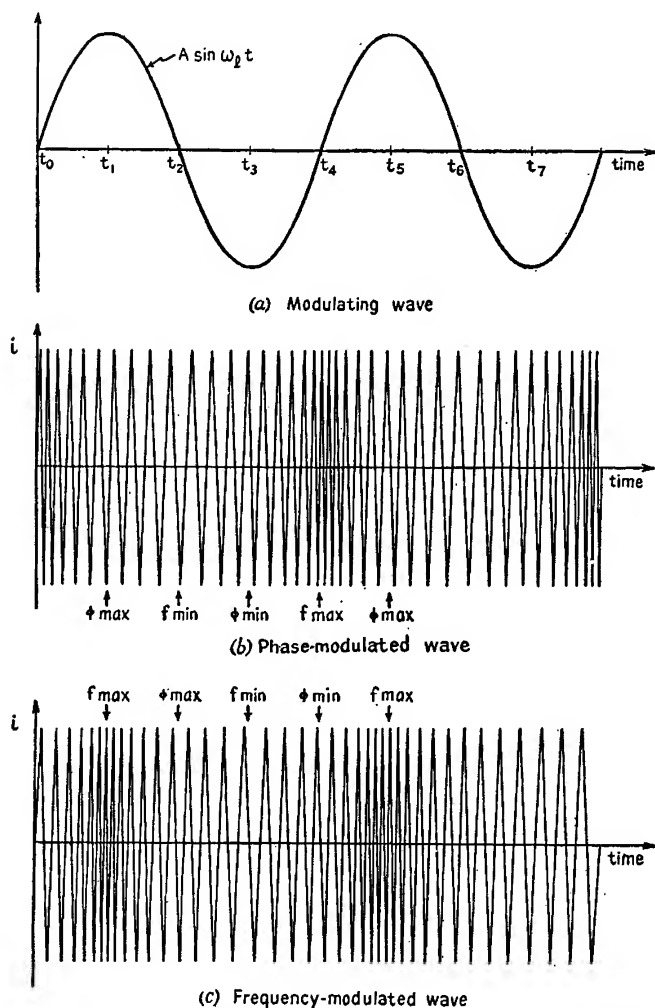


FIG. 7.2.—Phase-modulated and frequency-modulated waves resulting from sinusoidal modulation.

tional to the instantaneous value of the modulating wave, however great or small ϕ may be.

Analysis of (7.2) by Bessel functions shows that this current has many components, their number and relative amplitudes

depending upon the value of m_p . The relative amplitudes of the components having appreciable amplitudes are shown, in Fig. 7.3, for sinusoidal modulation with different values of maximum phase shift.¹ The components are spaced at intervals equal to the frequency of modulation f_i . As m_p increases, the number of appreciable side-frequency components increases.

The average power of a phase-modulated wave is equal to that of the unmodulated wave since the amplitude is unchanged. During phase modulation, energy is diverted from the center or carrier frequency into the side bands, not added to the wave as in amplitude modulation. If the amplitudes of all the components in the spectrum of an angle-modulated wave are squared, the sum of the squares equals the square of the amplitude of the unmodulated wave.

Shifts in phase cannot occur without changes in frequency. When the angle ϕ is increasing in the positive direction, the vector representing the phase-modulated wave is advancing with respect to the vector representing the unmodulated wave. An observer rotating with the "unmodulated" vector then would observe the "modulated" vector pulling ahead of him just as a speedier runner pulls ahead of a steady runner in a track race. While one runner is gaining, his speed is faster. Hence, when ϕ is increasing, the modulated vector is rotating faster, *i.e.*, at a higher frequency. When ϕ is stationary, as it is momentarily at the crest of a modulating signal, the speeds of the two vectors are the same, just as the speeds of two runners are the same when their relative position is unchanged. Hence when the phase-shift angle ϕ is increasing, the frequency of the modulated wave is higher than the carrier frequency; and when the phase-shift angle ϕ is decreasing, the frequency of the modulated wave is less than the carrier frequency.

That is, when the total angle of (6.1) or (7.1) is increasing at a greater than average rate, the frequency must be higher than the average frequency, which is f_c , the carrier frequency. When the total angle is increasing at a slower than average rate, the frequency

¹ A detailed discussion of the analysis of angle-modulated waves is found in "Frequency Modulation," pp. 1-55, 347-355, by August Hund, McGraw-Hill Book Company, Inc., 1942. There is an infinite number of side-band frequencies spaced f_i apart, but those which are considerably removed from the carrier frequency have negligible amplitudes. Only those having amplitudes equal to or greater than 1 per cent of the unmodulated carrier amplitude are shown here.

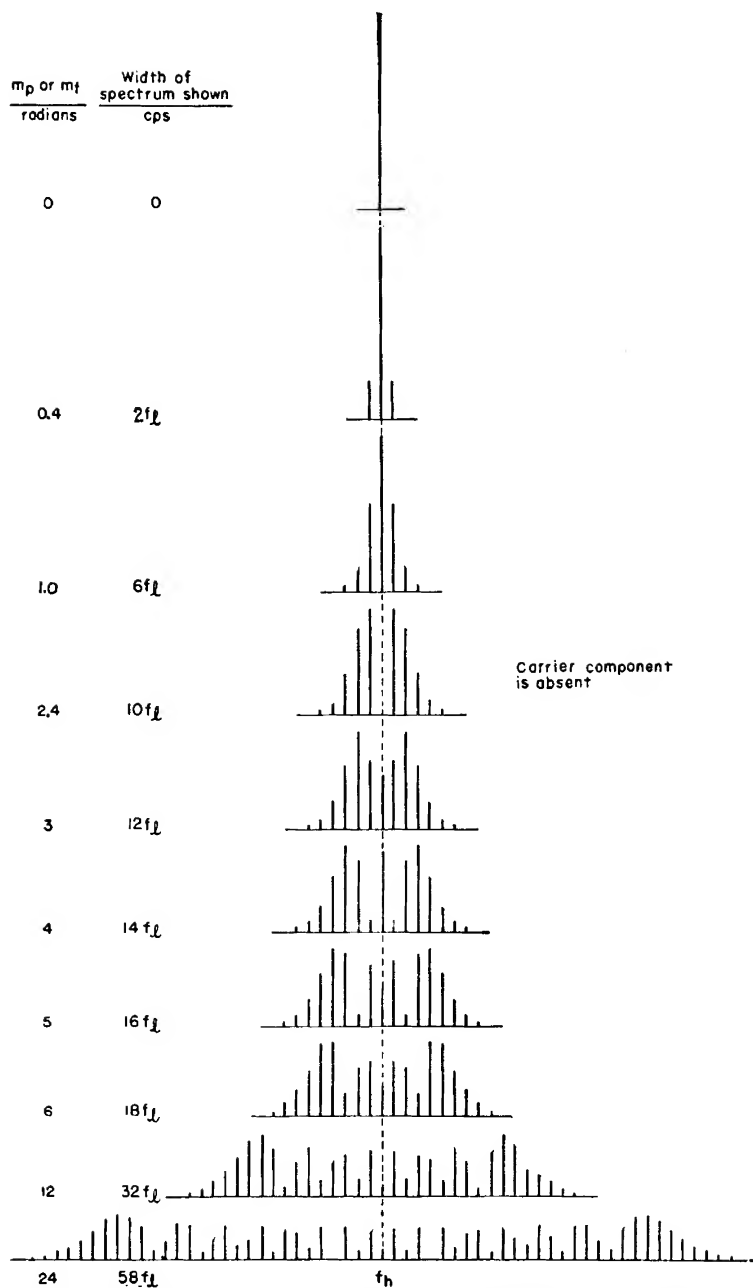


FIG. 7.3.—Amplitude spectra of the components of an angle-modulated wave resulting from sinusoidal modulation. The separation between components is f_L .

must be lower than f_h . The instantaneous frequency is given by

$$f = \frac{1}{2\pi} \frac{d\theta}{dt} \quad (7.3)$$

The instantaneous frequency during phase modulation by a sinusoidal modulating signal is, from (7.3) and (7.1),

$$f = f_h + m_p f_i \cos \omega_i t \quad (7.4)$$

The frequency is greatest when $\cos \omega_i t = +1$ and least when $\cos \omega_i t = -1$. The maximum shift in frequency with respect to the carrier frequency is $m_p f_i$. When $\cos \omega_i t = 0$, the wave has its unmodulated frequency. The changes in frequency that accompany sinusoidal phase modulation are cosinusoidal, *i.e.*, they lead the changes in phase by one-quarter of a modulation period. These variations in frequency are indicated in Fig. 7.2*b* by periodic changes in the distance between transitions of the curve across the time axis. The alternations of the wave are most compressed when the frequency is highest and least compressed when the frequency is lowest. Referring to Fig. 7.1, the angular velocity of the vector and the corresponding frequency are greatest during sinusoidal modulation when the vector representing the phase-modulated wave is advancing through the position *OP*, and are least when the vector is falling back through the position *OP*. At positions *OM* and *ON* the vector is momentarily stationary with respect to the position *OP*, and its angular velocity and frequency are momentarily the same as in the absence of modulation.

8. Frequency Modulation.—When there is no modulation, the representative vector, Fig. 1.1, rotates at constant speed and the angle θ of (6.1) increases uniformly at the rate ω_h radians/sec. During frequency modulation, as defined in Sec. 6, the *speed* of the rotating vector is alternately increased and decreased by amounts that are proportional to the successive positive and negative values of the modulating signal. If stop-motion pictures of the representative vector were taken regularly f_h times per second, in the absence of modulation the vector would be seen always in a fixed position, such as *OP* in Fig. 8.1. During frequency modulation the snapshots would show the vector advancing counterclockwise from the position *OP* during the time when the modulating signal is positive, because during this time the speed would be greater than f_h rps and the vector would rotate more than 1 revolution between snapshots. For the sinusoidal modulating signal of Fig. 7.2*a*, this advance in

position would occur during the time interval t_0 to t_2 . At time t_2 the speed is again f_h rps, the vector would rotate but 1 revolution between snapshots, and the snapshots would show no further advance in position. The position of maximum advance would appear as a position such as OM , Fig. 8.1. During the time interval t_2 to t_4 , Fig. 7.2*a*, the instantaneous value of the modulating signal is negative, and the speed of the rotating vector is less than f_h rps. Between snapshots the vector rotates less than 1 revolution, so that the vector would appear to move backward, or clockwise, retreating from the position of maximum advance. By the time t_4 , Fig. 7.2*a*, the snapshots would show that the vector had returned to its original position, and the cycle would repeat itself.

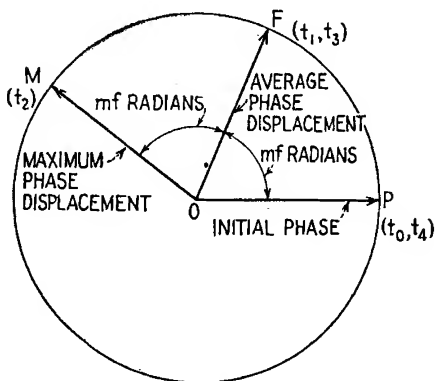


FIG. 8.1.—Motion of a vector representing a frequency-modulated current resulting from sinusoidal modulation, with f_h rps subtracted from its motion.

This effect may be visualized also by imagining two vectors, one rotating uniformly, representing an unmodulated wave, and the other rotating at the nonuniform speed, corresponding to frequency modulation. The observer may then imagine himself riding around on the steadily rotating unmodulated carrier-frequency vector, looking in the direction of rotation. When the instantaneous value of the modulating signal is positive, the speed of the frequency-modulated vector is faster than the unmodulated speed. The frequency-modulated vector would advance ahead of the observer, just as one runner pulls ahead of another when his speed is relatively increased. As long as the modulating signal is positive, the speed of the modulated vector is greater than the unmodulated speed and the modulated vector continues to pull ahead of the steady-running unmodulated vector. When the modulating signal is

negative, the speed of the frequency-modulated vector is less than the speed of the steady-running unmodulated vector. The slow-down causes the frequency-modulated vector to lose its lead, just as a runner who has attained an advanced position will decrease his lead if he drops his speed below that of his steady-running rival. In frequency modulation the speed-up during the positive half cycle of the modulating signal is matched by a corresponding slowdown during the negative half cycle of the modulating signal; there is no net gain over the complete cycle. Hence the rotating observer sees the frequency-modulated vector get ahead of him during the first, or positive, half cycle of the modulating signal and then lose this advance and return to his own position at the end of the second, or negative, half cycle of the modulating signal. Figure 8.1 illustrates this change in relative position, and Fig. 7.2c shows the corresponding variations in the instantaneous current. The changes in frequency are indicated in Fig. 7.2c by the periodic changes in the distance between transitions of the curve across the time axis.

From this discussion and from Figs. 7.1, 7.2, 8.1, it is clear that there are certain similarities and certain differences between phase and frequency modulation. It is appropriate at this point to repeat the statements made in the last paragraph of Sec. 6. (1) The sine in (6.1) involves an *angle*, not a speed. (2) The total angle θ is not equal to ωt when ω is not a constant. (3) Frequency is proportional to the *rate of change* of the total angle θ of (6.1).

Let the maximum frequency change (or frequency deviation, as it is also called) occurring in the frequency modulation indicated in Fig. 7.2c be denoted by Δf . The instantaneous frequency change is proportional to the instantaneous value of the sinusoidal modulating signal. Then the actual frequency at any instant t is given by

$$f = f_h + \Delta f \sin \omega t \quad (8.1)$$

The expression for the instantaneous frequency in terms of the angular displacement is given by (7.3). From (7.3) and (8.1),

$$f = \frac{1}{2\pi} \frac{d\theta}{dt} = f_h + \Delta f \sin \omega t \quad (8.2)$$

Then

$$d\theta = 2\pi(f_h + \Delta f \sin \omega t) dt$$

Integrating to obtain θ ,

$$\theta = 2\pi f_h t - \frac{\Delta f}{f_l} \cos \omega t + (\text{a constant}) \quad (8.3)$$

The constant is evaluated from the condition that $\theta = 0$ at $t = 0$. The constant equals $\Delta f/f_i$, and (8.3) becomes

$$\theta = \omega_h t + \frac{\Delta f}{f_i} (1 - \cos \omega_i t) \quad (8.4)$$

The second term on the right of (8.4) is the departure of the angle θ of the frequency-modulated wave from the value it would have in the absence of modulation. It is the angle observed by the rotating observer of the previous discussion. The maximum departure is $2 \Delta f/f_i$, and the average departure is $\Delta f/f_i$. With a sinusoidal modulating signal, there is a cosinusoidal departure about the average with an amplitude, or peak value, $\Delta f/f_i$. This amplitude, or maximum departure from the average during frequency modulation, is denoted by the symbol m_f and is called the *modulation index* for frequency modulation.¹

$$m_f = \frac{\Delta f}{f_i} \text{ radians} \quad (8.5)$$

This is really an angle in radians, similar to the angle designated by m_p (radians) in the discussion of phase modulation, Sec. 7. From (8.5), (8.4) becomes

$$\theta = \omega_h t + m_f (1 - \cos \omega_i t) \quad (8.6)$$

and (6.1) becomes

$$i = \hat{I} \sin [\omega_h t + m_f (1 - \cos \omega_i t)] \quad (8.7)$$

This is the final expression for a frequency-modulated current resulting from modulation by a sinusoidal modulating signal, as shown in Fig. 7.2c. Notice how this expression differs from (7.2), which is for a phase-modulated current.

In principle, there is no limit to Δf for distortion-free modulation. As long as the frequency deviation is proportional to the instantaneous value of the modulating signal, the modulation is free of distortion no matter how great or how small the frequency deviation may be.

Because of the similarity of Figs. 7.2b and c, sinusoidal phase and frequency modulation have identical spectra when the maximum angular departure from the average is the same for both, i.e., when $m_f = m_p$. Thus the spectra of Fig. 7.3 apply to both frequency modulation and phase modulation. It follows that,

¹ It is also called the "deviation ratio."

in angle-modulation schemes, power is diverted from the carrier into the side bands, the total power remaining unchanged. For certain values of m_f , such as 2.40, 5.52, 8.65, . . . , all the power is diverted into the side bands and the carrier component disappears. This fact is used in measuring the frequency deviation of frequency-modulated waves.

With frequency modulation, doubling the frequency without changing the amplitude of the sinusoidal modulating signal halves the value of m_f , (8.5), since f_i is doubled while Δf remains unchanged. This approximately halves the number of spectrum lines, Fig. 7.3. The new lines, however, are twice as far apart in frequency; hence the spectrum width is not greatly changed. This is a sharp contrast to phase modulation, in which doubling the value of f_i does not change the value of m_p or the number of lines, but places the lines twice as far apart, thus doubling the width of the spectrum.

In Fig. 8.2 is shown the amplitude spectrum of a sinusoidally *frequency-modulated* wave whose *frequency* deviation is 12,000 cps. Only those side frequencies are shown whose amplitude is 1 per cent or more of the amplitude of the unmodulated carrier. When the frequency of the modulated signal is 1,000 cps, Fig. 8.2a, the number of side frequencies on either side of the carrier is 16, spaced 1,000 cps apart. The total band width is 32,000 cps. The band width is defined here as the minimum band width containing all components of an amplitude equal to or greater than 1 per cent of the amplitude of the unmodulated carrier.

If the frequency of the modulating signal is doubled to 2,000 cps, Fig. 8.2b, the frequency deviation being kept the same, the number of side frequencies on either side of the carrier decreases to 9, spaced 2,000 cps apart. The total band width is 36,000 cps. With the frequency of the modulating signal again doubled to 4,000 cps, the number of side frequencies decreases to 6, but they are spaced 4,000 cps apart so that the band width is 48,000 cps. As the frequency of the modulating signal increases, the frequency deviation remaining the same, the modulation index m_f decreases, indicating that the maximum *angular* departure from the average decreases.

Figure 8.3 shows the amplitude spectrum for a sinusoidally *phase-modulated* wave, with the maximum angular departure chosen so that the spectrum at the modulating frequency of 1,000 cps, Fig. 8.3a, is identical with that of Fig. 8.2a. In phase modulation, m_p is constant if the amplitude of the modulating signal is constant.

FREQUENCY MODULATION

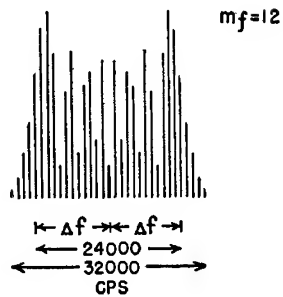
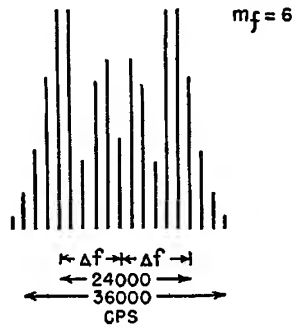
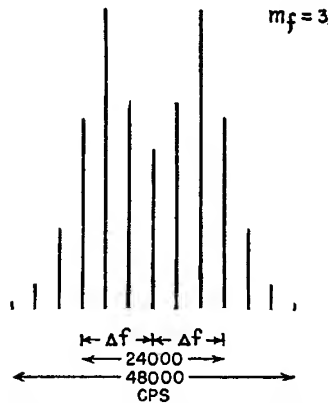
(a) $f_c = 1000$ CPS(b) $f_c = 2000$ CPS(c) $f_c = 4000$ CPS

FIG. 8.2.—Amplitude spectrum for a frequency-modulated wave as the modulation frequency f_i increases; Δf is constant at 12,000 cps.

PHASE MODULATION

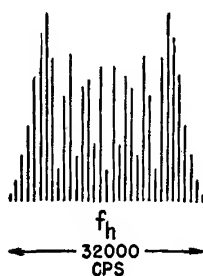
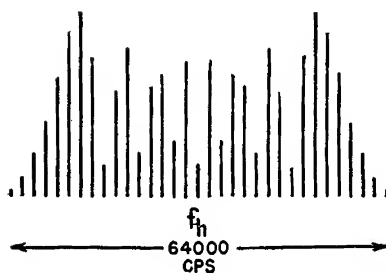
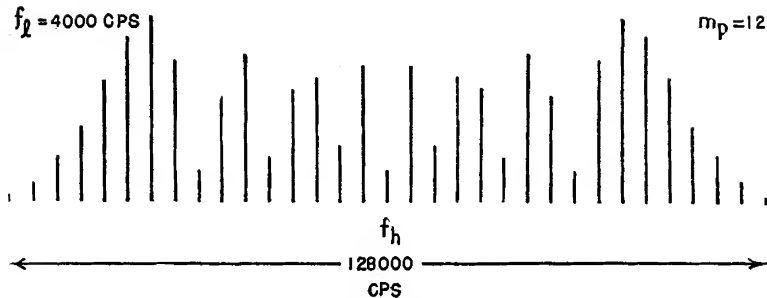
(a) $f_l = 1000$ CPS $m_p = 12$ (b) $f_l = 2000$ CPS $m_p = 12$ (c) $f_l = 4000$ CPS $m_p = 12$ 

FIG. 8.3.—Amplitude spectrum for a phase-modulated wave as the modulation frequency f_l increases; m_p is constant and chosen to yield at $f_l = 1,000$ cps the same spectrum as in Fig. 8.2a.

Therefore, when a modulating signal of constant amplitude has its frequency doubled, m_p remains the same and the number of spectrum lines remains the same, but their spacing is doubled and the band width is doubled. Hence in this case the number of spectrum lines on either side of the carrier remains constant at 16, and the band width increases in direct proportion to the frequency of the modulating signal, becoming 128,000 cps, Fig. 8.3c, at a modulating frequency of 4,000 cps.

The question often arises as to the ratio of the band width (required for all the significant components of a frequency-modu-

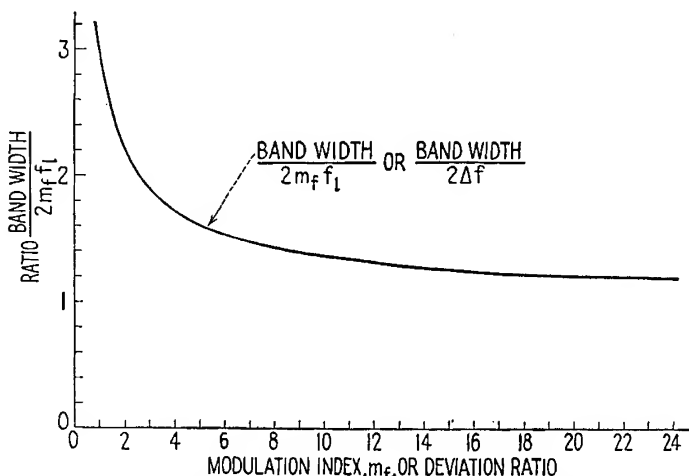


FIG. 8.4.—Curve showing the ratio of the band width to twice the frequency deviation for sinusoidal phase or frequency modulation.

lated wave) to the frequency range over which the *instantaneous* frequency swings during modulation. The band width is always greater than the range $2\Delta f$ swept through by the instantaneous frequency during modulation. In Fig. 8.4 is plotted, for various values of m_f , the ratio of the band width to the range $2\Delta f$ over which the instantaneous frequency varies. The figure is obtained by dividing the band width as above defined by $2\Delta f$,¹ plotting this ratio as a function of m_f , and drawing a smooth curve from the points so obtained. For a modulation index of 5, the ratio of band width to $2\Delta f$ is 1.6. This means, for example, that if the frequency deviation Δf is 15 keps and the modulating frequency is 3 keps, the band width is 1.6 times $2\Delta f$ or 48 keps, not 30 keps, which is the

¹ Or by dividing the number of significant components on either side of the carrier by m_f .

frequency range over which the instantaneous frequency is varied. The greater the modulation index, the smaller the difference between the band width and $2 \Delta f$.

From Figs. 8.2 and 8.3 it is seen that the band width remains more nearly constant with frequency modulation than with phase modulation, so that frequency modulation utilizes a given channel much more effectively than phase modulation. A channel width of 200 keps is ample for high-fidelity wide-band frequency-modulation (FM) broadcasting, in which 75 keps is the maximum frequency deviation. A channel width of the order of 30 or 40 keps is sufficient for police and similar "frequency-modulation" transmission, where the maximum frequency deviation is smaller.

The advantages of frequency modulation as compared with amplitude modulation are less interference, less noise compared to the desired receiver output, less cross interference between neighboring stations, and better fidelity of reproduction for the same expenditure of power. The frequency-modulation systems minimize interference due to static and noise by employing a receiver whose output depends only upon changes in frequency, any changes in amplitude being removed by a limiter stage or stages. By utilizing large changes to convey the message, the changes caused by noise or by static can be made relatively very small. To secure the advantages of frequency modulation, the received signal strength must be great enough and the receiver so designed that the receiver output depends only upon the changes in frequency. The advantages of frequency modulation are not so pronounced when the deviation in phase or in frequency due to modulation is small. A disadvantage of frequency modulation compared with amplitude modulation is the greater band width required for the same fidelity of reproduction.

9. Frequency Modulation from Phase Modulation.—In Fig. 8.3 it is shown that in phase modulation if m_p , or the maximum angular departure, is maintained constant, the band width increases in direct proportion to the frequency. Now if the maximum angle m_p for Fig. 8.3b were reduced by a factor of 2 to the value of m_f for Fig. 8.2b, the spectra would agree again. This could be done by reducing the amplitude of the modulating signal by a factor of 2, since in phase modulation the maximum angular departure is proportional to the amplitude of the modulating signal. Similarly, if the maximum angle m_p for Fig. 8.3c were reduced by a factor of 4, the conditions of Fig. 8.2c would result. This may be obtained by reducing

the amplitude of the signal causing phase modulation by a factor of 4. Hence, if a network that reduces the amplitude of the modulating signal in direct proportion to the frequency of the signal precedes a device producing phase modulation, the net result is a frequency-modulated wave. Such a network is approximated, for example, by a large resistance in series with a large capacitance, Fig. 9.1.

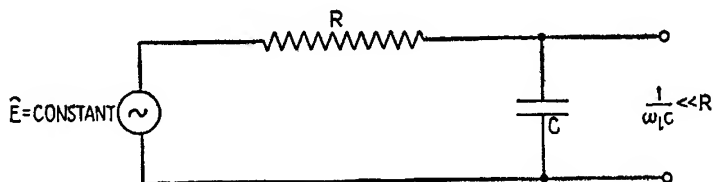


FIG. 9.1.—Network whose voltage output varies inversely as the impressed frequency.

Assuming the impedance connected to the output to be infinite, the amplitude of the output voltage varies inversely as the frequency for a constant input voltage. Such a circuit is called an integrating circuit¹ in other applications.

The principles of this method are illustrated by the waveforms in Figs. 9.2 and 9.3. Let the modulating signal be a square wave

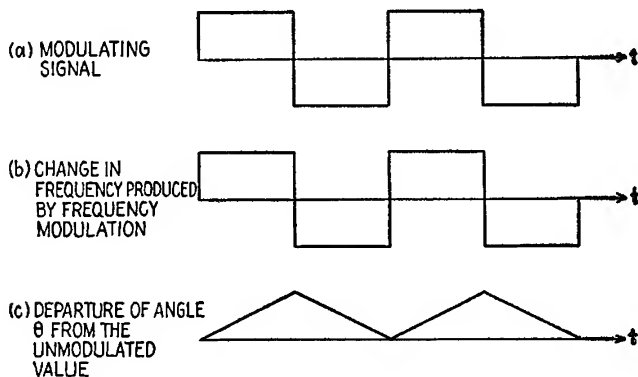


FIG. 9.2.—Waveforms in frequency modulation resulting from a square-wave modulating signal.

as in Fig. 9.2a. The corresponding frequency change due to frequency modulation is shown in Fig. 9.2b. The departure of the angle θ of (6.1) from the value it would have in the absence of modulation is shown in Fig. 9.2c. If the square-wave modulating signal is fed into an integrating circuit, the output has the waveform shown in Fig. 9.3b. The departure of the angle θ due to phase modulation

¹ See Chap. XXIV, Sec. 22, and also Chap. IX, Sec. 13.

tion has the same waveform, indicated in Fig. 9.3c. The corresponding change in frequency has the waveform¹ of Fig. 9.3d, which is that required for frequency modulation.

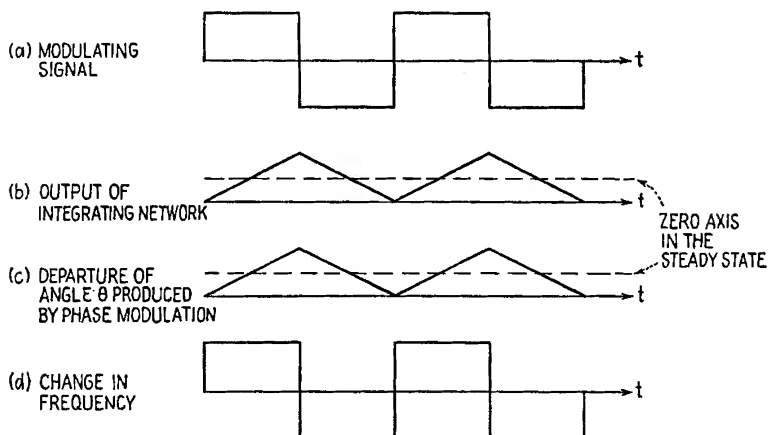


FIG. 9.3.—Waveforms in frequency modulation by means of integration, or imparting a $1/f$ characteristic, and phase modulation for a square-wave modulating signal.

10. Alteration of Modulated Wave by a Correctly Tuned Circuit.

Any type of modulated wave may be altered in passing through a tuned circuit, even though the circuit is tuned properly to the center component. The admittance of a single series LC circuit and the phase of its current with respect to an impressed sinusoidal emf are shown in Fig. 10.1, as functions of frequency. When a modulated voltage is impressed upon a correctly tuned circuit, the center or carrier component of the emf falls at the frequency of maximum admittance. The side components of the emf necessarily fall where the circuit is less sensitive. The side-frequency currents are weakened relative to the current at the center frequency and also are shifted in phase with respect to the corresponding side-frequency emf's.

For example, assume an impressed emf that is amplitude-modulated 100 per cent and is described by the equation

$$e = \hat{E} \sin \omega_h t + \frac{1}{2} \hat{E} \cos (\omega_h - \omega_l) t - \frac{1}{2} \hat{E} \cos (\omega_h + \omega_l) t \quad (10.1)$$

¹ In Fig. 9.3b the response of the network is approximately as indicated for the first few cycles. In the steady state there is no constant component, and the triangular waveform is centered about the zero axis (as in Fig. 13.1, Chap. IX). However, in the steady state the frequency changes are as indicated here in Fig. 9.3d.

Assume, further, that the circuit is properly tuned, *i.e.*, to the carrier frequency, and that the side-frequency components fall at the half-power frequencies of the circuit. These conditions are shown by the amplitude spectrum superimposed upon the frequency axis in Fig. 10.1. (In this figure, the dissymmetry of the resonance curve, which is slight in the case of a high- Q circuit, is neglected). The admittance at the half-power frequencies is 70.7 per cent as great as the admittance at resonance, and in a series circuit has a phase angle of $+45^\circ$ at the lower half-power frequency and -45° at

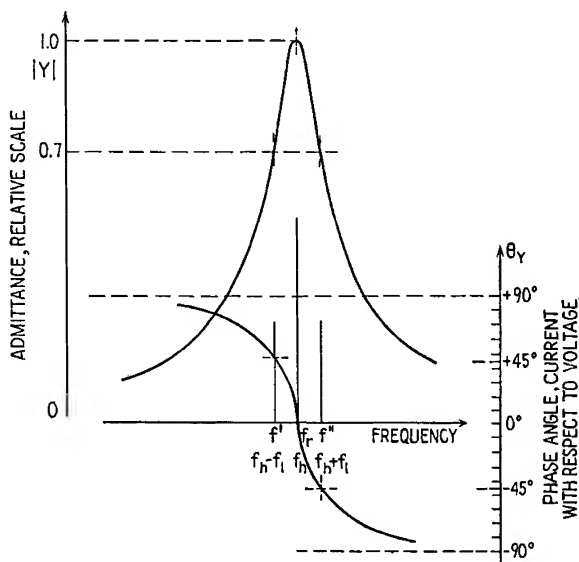


FIG. 10.1.—Admittance and phase-angle curves for a high- Q LC circuit, with an amplitude spectrum of the induced voltage of (10.1) shown on the frequency axis.

the upper half-power frequency. Under the conditions here assumed, the side-frequency currents are weakened with respect to the carrier current by a factor of 0.707; the lower side-frequency current is shifted ahead 45° with respect to the lower side-frequency emf; and the upper side-frequency current is shifted 45° behind its emf.

Figure 10.2*a* shows the vectors corresponding to the emf's of (10.1) at the time $t = 0$, and Fig. 10.2*b* shows the vectors representing the currents caused by these emf's. The side-frequency-current vectors are relatively shortened in length and shifted in phase with respect to the corresponding emf vectors, in accordance

with the admittances explained above. When the total current is obtained by a vector addition similar to that of Fig. 5.1, it is found that the total-current vector remains parallel with the carrier-component vector. Complete absence of angle modulation thus is indicated. The length of the total-current vector is shown as a function of time in Fig. 10.3, and also the length of the total-voltage vector. The variation in length of the total-current vector is such that the current is only 70.7 per cent modulated, and the modulation envelope is retarded one-eighth of a modulation period, as shown in Fig. 10.3c.

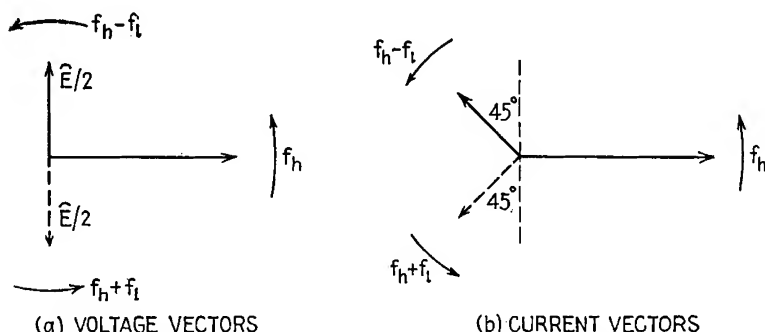


FIG. 10.2.—Vectors representing the component voltages and currents at $t = 0$, in an LC circuit with the voltage expressed by (10.1) under the conditions of Fig. 10.1.

In general, a sinusoidally amplitude-modulated voltage applied to a correctly tuned parallel-resonant circuit causes a total current having a reduced degree of modulation and a retarded envelope. The closer the two side frequencies lie to the center frequency, the smaller these effects will be.

Side frequencies that fall in the neighborhood of or beyond the half-power frequencies are considerably attenuated. This effect is known as side-band cutting, or clipping. Attenuation and retardation of the higher frequency components of a complex modulating wave through side-band cutting in amplitude-modulation systems constitute one form of signal distortion. When a circuit is to handle a modulated wave, the Q of the circuit is an important factor, because of the relationship that exists between Q , and the half-power band width. When the band width of the spectrum of the modulated wave is wide, low- Q resonant circuits or band-pass filter circuits must be employed.

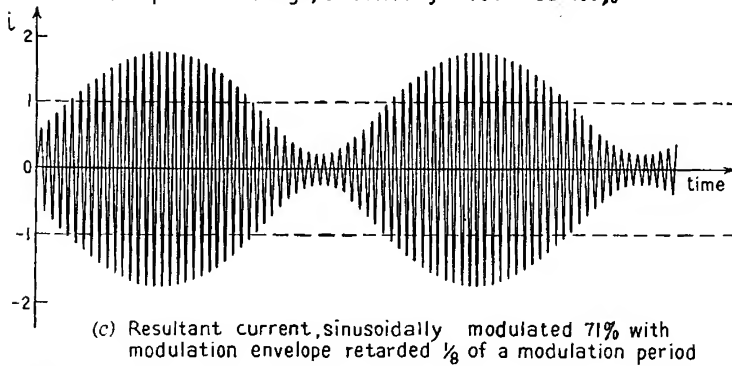
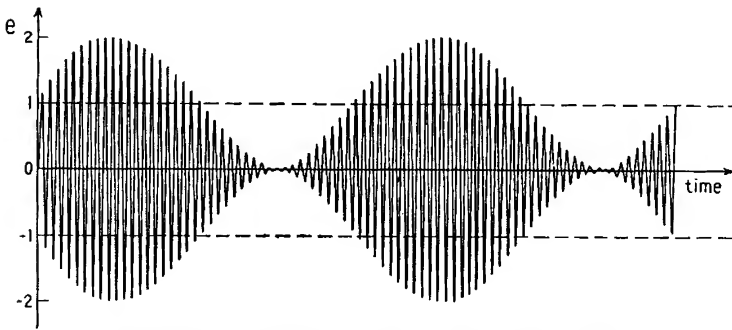
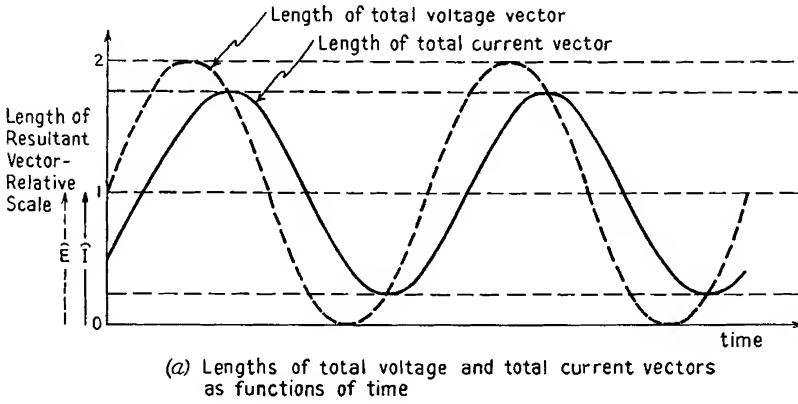


FIG. 10.3.—Alteration of an amplitude-modulated wave by a properly tuned circuit

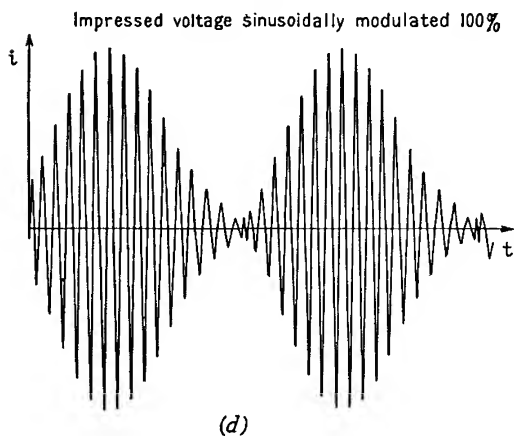
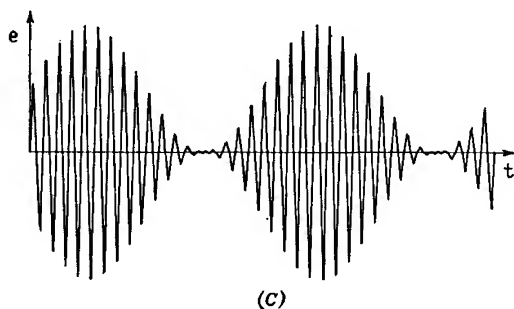
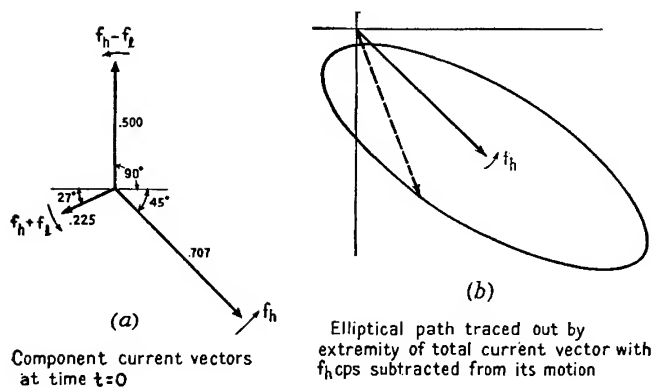


FIG. 11.1.—Distortion of an amplitude-modulated wave by an off-tune circuit under conditions explained in Sec. 11 of the text.

11. Distortion of a Modulated Wave by an Off-tune Circuit.—

A circuit that is detuned from the center frequency of a modulated wave causes distortion of the wave. To illustrate by an example, suppose that the emf of (10.1) is applied to a series LC circuit tuned to the frequency of the lower side component and that the frequency of the carrier coincides with the upper half-power frequency of the circuit. Vectors that represent the components of the current then have the relative lengths shown in Fig. 11.1*a* and at $t = 0$ are in the positions indicated.

With the frequency f_h subtracted from the rotational motion of the total-current vector, the outer extremity of this vector traces out an ellipse, Fig. 11.1*b*, once per modulation period. The vector changes in length and moves back and forth from one side to the other of the carrier vector. The current in the circuit is non-sinusoidally amplitude-modulated and also nonsinusoidally angle-modulated. The distorted current is shown as a function of time in Fig. 11.1*d*.

12. Distortion of a Modulated Wave from Other Causes.—Any influence that alters the relative phases of the components of a modulated wave or alters their relative amplitudes alters the modulation. The alteration may occur in the transmitter, in the medium between the transmitter and the receiver, or in the receiver.

In the transmitter, the circuits should be tuned to the center frequency and should have a Q sufficiently low so as not to attenuate the sidebands. If the components of a modulated wave are propagated between transmitter and receiver with unequal attenuation, unequal time delay, or both, the waveform at the receiver differs from that at the transmitter. This type of distortion may be caused by nonlinear effects in the ionosphere or by interference resulting from simultaneous propagation over two or more paths.¹ The problem of minimizing the distortion in the r-f circuits of the receiver is similar to the same problem in respect to the transmitter.

13. Spectra of Typical Signals.—The spectrum of a radio signal presents information that is useful to the engineer in several ways. A spectrum may show the *amplitude* of each frequency component plotted vertically against a horizontal scale of frequency. Or, following the customary procedure in optics, the spectrum may show the *energy* of the components, or the square of their ampli-

¹ R. W. P. KING, H. R. MIMNO, and A. H. WING, "Transmission Lines, Antennas, and Wave Guides," Chap. IV on Wave Propagation, McGraw-Hill Book Company, Inc., 1945.

tudes as a function of frequency. The indicated distribution of amplitude or of energy may persist continuously, or the distribution may vary with time, in which case a time-average distribution may be of interest, such as the one shown in Fig. 4.2.

Figure 4.2 was designed to show that, with amplitude modulation, the radio signal requires a channel whose width is twice the frequency of the highest frequency component in a complex modulating signal. A similar time-average spectrum of a wide-swing frequency-modulated wave would be approximately 200 keps wide, while that of an amplitude-modulated television signal would be 4 mcps or more in width.

The spectrum of a continuous sinusoidal wave consists of a single fine line, Fig. 13.1a. The spectrum of a continuous wave whose amplitude varies sinusoidally consists of three fine lines, Fig. 13.1b. If the wave is 100 per cent modulated, the outer lines of its *energy* spectrum are one-fourth as tall as the middle line. The energy spectrum of a continuous wave that is either frequency-modulated or phase-modulated sinusoidally would resemble those of Fig. 7.3 after all vertical distances shown there are squared.

Signals consisting of pulses or groups of oscillations may be analyzed and the spectrum determined by Fourier analysis. The spectrum of a group of oscillations of constant amplitude and frequency is *not* a single line of infinitesimal width. The spectrum occupies a finite distance along the frequency axis, Fig. 13.1c. The *fewer* the number of oscillations in the group, the *broad*er the spectrum. There may be some minima and subsidiary maxima in the spectrum, as is shown in the figure.

The spectrum of a group of oscillations of fixed frequency but of decreasing amplitude is shown in Fig. 13.1d. The greater the decrement (rate of decrease in amplitude), the broader the spectrum. Its highest point is located close to the frequency of the oscillations. "Spark" transmitters of the early days of radio radiated such trains of damped oscillations.

If the decrement of a damped wave train is so great that only one or two oscillations of appreciable intensity exist in any one group, the spectrum is so broad as to spread over much of the used radio frequencies, Fig. 13.1e. Disturbances due to electric discharges in the atmosphere are of this general shape.

A pulse that consists of a single half cycle of a sinusoid, Fig. 13.1f, has a spectrum that spreads over all frequencies. The height of the spectrum is greatest at zero frequency. At the fre-

quency corresponding to that of the sinusoid the energy spectrum has a height of 0.62 of the height at zero frequency. The shorter the duration of the pulse, the more uniform the distribution of energy in the radio spectrum. If the pulse is of negligible width, its spectrum has uniform height, Fig. 13.1*g*.

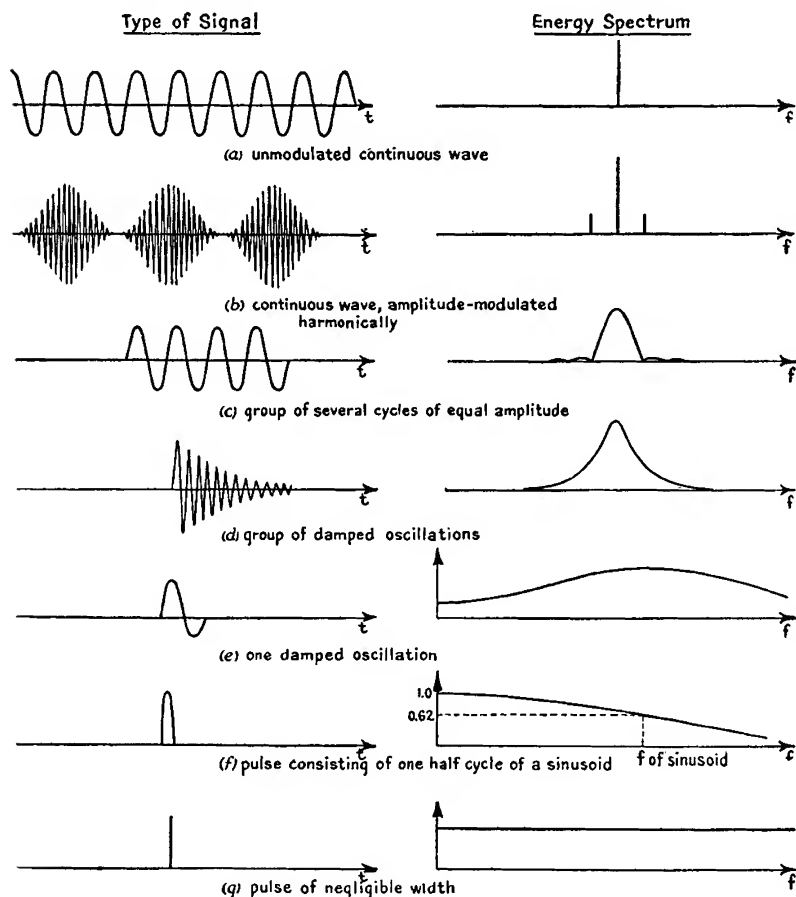


FIG. 13.1.—Various types of signals and their energy spectra.

Although the signal shown in Fig. 13.1*a* would be completely admitted by a circuit having a very high Q , the signals shown in *b*, *c*, *d* would be completely admitted only if the circuit has an admittance curve that spans the spectrum of the signal. On the other hand, signals of the types shown in *e*, *f*, *g* have such broad spectra that they will cause some disturbances in any circuit.

CHAPTER XX

METHODS OF MODULATION

I. METHODS OF AMPLITUDE MODULATION

1. **General.**—Amplitude modulation may be produced by many methods and devices. Those which have been used commonly are

1. Absorption processes:
 - a. Microphone in antenna system
 - b. White system
 - c. Magnetic amplifier
2. Modulated Class *C* amplifier:
 - a. Plate-modulated
 - b. Grid-modulated
 - c. Cathode-modulated
3. Amplitude-modulated oscillator
4. Square-law processes:
 - a. Nonlinear circuit element
 - b. Triode square-law modulator
 - c. Balanced square-law modulator

Modulation processes based upon the absorption of r-f energy were important in the early days of radio. A microphone placed in series with an antenna absorbed more or less r-f energy as the resistance of the microphone varied, and in consequence the amplitude of the antenna current was altered. The White system produced a similar effect by means of a vacuum tube whose plate was connected to one end of the antenna coil and whose cathode was connected to the other end. When a low-frequency voltage was applied to the grid, the plate-to-cathode resistance of the tube was altered and thus a varying amount of r-f energy was absorbed. The magnetic-amplifier method made use of an iron-core reactor in series with the antenna. A unidirectional current in another winding on the core was varied at an audio-frequency rate. The varying direct current caused the iron core to become more or less saturated, causing the inductance of the reactor to decrease and increase, resulting finally in variation of the amplitude of the antenna current.

Nowadays an amplitude-modulated wave is produced usually by altering one or more of the polarizing or supply voltages in an r-f amplifier or oscillator in accordance with the modulating signal wave. The variation in the amplitude of the r-f output of the amplifier or oscillator is proportional to the successive values of the modulating signal if the device is properly adjusted.

2. Plate-modulated Class C Amplifier.—A Class C amplifier can be designed and operated so that the amplitude of the r-f output is directly proportional to the plate-supply voltage. An amplitude-modulated wave may be produced by varying the plate-supply voltage of such an amplifier in accordance with a low-frequency modulating signal.

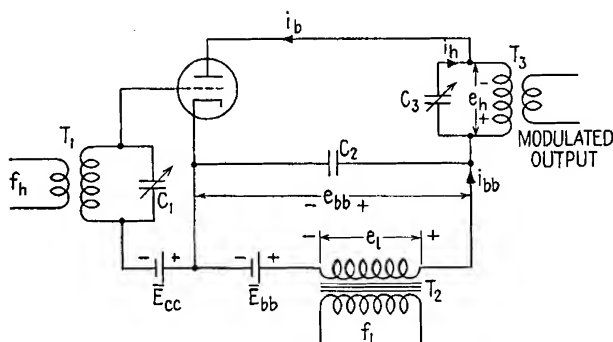


FIG. 2.1.—Basic diagram of a plate-modulated Class C amplifier.

A plate-modulated Class C amplifier circuit is shown in Fig. 2.1.¹ Capacitor C_1 tunes the grid circuit to the carrier frequency, and C_3 tunes the plate circuit to the same frequency. The total instantaneous plate-supply voltage e_{bb} is the sum of the fixed voltage \bar{E}_{bb} and the secondary voltage of T_2 , which is an iron-core modulation transformer. Capacitor C_2 serves as an r-f by-pass, to keep the lower side of the plate tank circuit at the same r-f potential as the cathode.

The grid circuit is excited at the carrier frequency by the previous stage in the transmitter. The alternating voltage developed across the grid tank circuit is in series with the fixed grid voltage, \bar{E}_{ce} , whose magnitude is two to four times the value that would reduce the plate current to zero in the absence of r-f excita-

¹ For simplicity, a neutralization arrangement to prevent self-oscillation that might occur in such a triode amplifier has been omitted from the diagram.

tion and low-frequency modulating signal. The potential of the grid varies in the manner indicated in Fig. 2.2*a*. A pulse of plate current occurs once during each r-f cycle during the interval when the instantaneous grid voltage is more positive than the cutoff value (ideally, above e_b/μ), Fig. 2.2*c*. These pulses draw their power from the plate-supply source. The plate circuit is tuned to the frequency of these pulses, *i.e.*, to the carrier frequency, and the pulses set the plate tank circuit into oscillation. In doing so they deliver power to the tank circuit. The Class *C* amplifier is thus a power-converting device¹ that takes power from the low-frequency plate supply and converts part of it into r-f power in the tank circuit. The part that is not converted into r-f power is dissipated, mostly at the plate of the tube.

When there is no modulation, the plate-supply voltage is from the d-c source alone, Fig. 2.1, so that the plate-supply voltage e_{bb} , Fig. 2.2*b*, has the steady value \bar{E}_{bb} . The pulses of plate current due to the r-f voltage applied to the grid are all alike, occurring at the carrier frequency. The plate tank circuit, tuned to this frequency, is excited into steady oscillation. The pulsating non-sinusoidal plate current contains a d-c component, a component of fundamental, or carrier, frequency, and harmonics of the carrier frequency.² The resistance of the tank coil to direct current is usually negligible, and owing to resonance the impedance of the tank circuit, considered as a parallel-resonant circuit, is very small at the harmonic frequencies. Since the tank circuit has a large impedance at the carrier frequency, the voltage across the tank circuit is due principally to the fundamental- or carrier-frequency component of the plate current and is practically sinusoidal in waveform. This voltage is indicated as e_a in Fig. 2.1. Since the voltage across the tank circuit is practically sinusoidal, the current in either branch is also practically sinusoidal in waveform.

During modulation, the total instantaneous plate-supply voltage, e_{bb} , Figs. 2.1 and 2.2*b*, swings above and below the value \bar{E}_{bb} because of the modulating signal voltage e_i appearing across the secondary of transformer T_2 , Fig. 2.1. As a result, the pulses of plate current flow during a greater or smaller portion of the r-f cycle and increase and decrease in magnitude, Fig. 2.2*c*. With proper adjustment and operation of the modulated Class *C* amplifier, the changes in the size and shape of the pulses cause the amplitude of

¹ A more detailed treatment of the Class *C* amplifier is given in Chap. XIV.

² See Chap. IX.

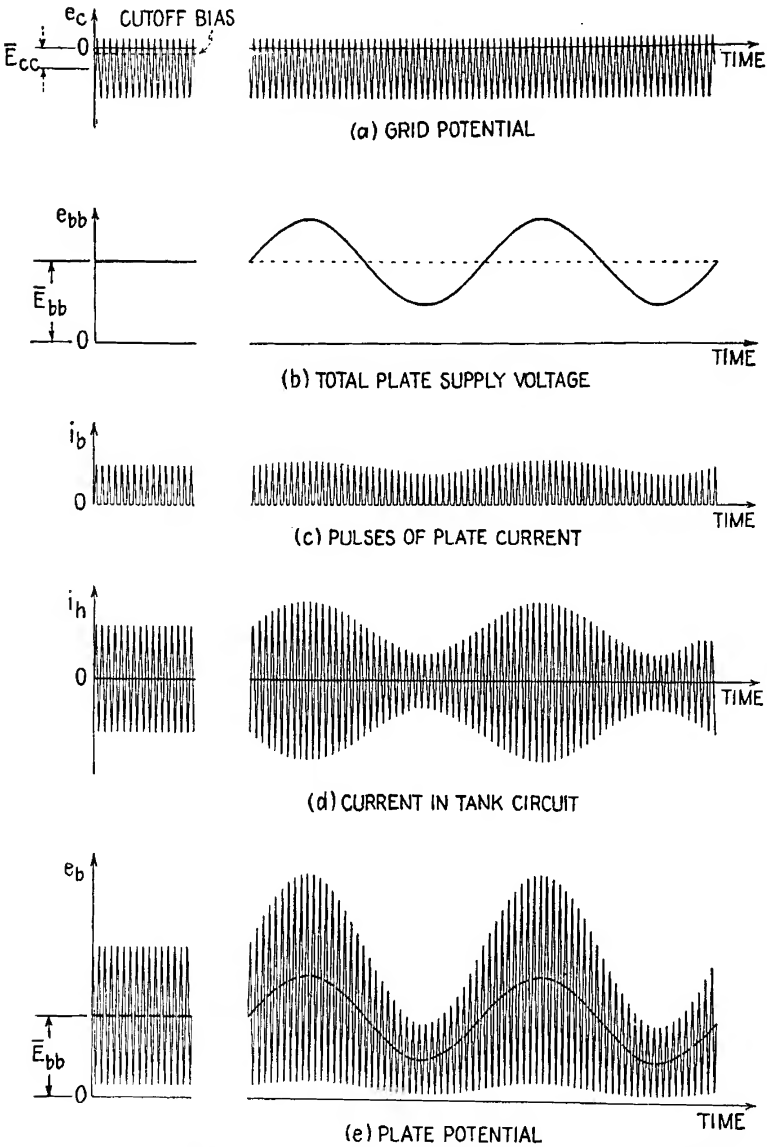


FIG. 2.2.—Voltages and currents in a plate-modulated Class C amplifier.

the r-f current in the tank circuit to vary in proportion to the total plate-supply voltage.

The total plate-supply voltage with a sinusoidal modulating signal is

$$e_{bb} = \bar{E}_{bb} + \hat{E}_l \sin \omega_l t \quad (2.1)$$

where the subscript l as in Chap. XIX denotes the low, or modulating, frequency. The proportionality between the amplitude of the r-f current and the plate-supply voltage may be expressed as

$$\hat{I}_h = k e_{bb} \quad (2.2)$$

where k is the constant of proportionality. Substituting (2.1) into (2.2),

$$\hat{I}_h = k(\bar{E}_{bb} + \hat{E}_l \sin \omega_l t)$$

which may be rearranged to give

$$\hat{I}_h = k\bar{E}_{bb} \left(1 + \frac{\hat{E}_l}{\bar{E}_{bb}} \sin \omega_l t \right) \quad (2.3)$$

The variation in amplitude is the same as that of (2.1), Chap. XIX, and is the proper variation for an amplitude-modulated wave. The factor \hat{E}_l/\bar{E}_{bb} appears here in place of m_a in (2.1), Chap. XIX. Therefore, in this method of producing amplitude modulation, the degree of modulation is determined by the ratio of the modulating voltage to the d-c plate-supply voltage. For sinusoidal modulation,

$$m_a = \frac{\hat{E}_l}{\bar{E}_{bb}} \quad (2.4)$$

With this substitution (2.3) becomes

$$\hat{I}_h = k\bar{E}_{bb}(1 + m_a \sin \omega_l t) \quad (2.5)$$

and the equation for the *instantaneous* r-f current $i_h = \hat{I}_h \sin \omega_h t$ is

$$i_h = k\bar{E}_{bb}(1 + m_a \sin \omega_l t) \sin \omega_h t \quad (2.6)$$

which is the same in form as (2.1), Chap. XIX.

If the impedance of the tank circuit is essentially the same at the carrier and side-band frequencies, the amplitude of the r-f voltage across the tank circuit varies in the same manner as (2.5). The r-f voltage across the tank circuit is also proportional to the total plate-supply voltage, or

$$\hat{E}_h = k'\bar{E}_{bb}(1 + m_a \sin \omega_l t) \quad (2.7)$$

where k' is the constant of proportionality. From (2.4) it is seen that, when the amplitude of the secondary voltage of T_2 , Fig. 2.1, equals the fixed part of the plate-supply voltage, 100 per cent modulation is obtained. If \hat{E}_l exceeds \bar{E}_{bb} , the total plate-supply voltage e_{bb} goes negative during a part of each modulation period and overmodulation results. While e_{bb} is negative, there can be no pulses of plate current to excite oscillations in the tank circuit. Assuming \hat{I}_h to be proportional to e_{bb} while e_{bb} is positive, the overmodulated tank-circuit current then has the waveform illustrated in Fig. 2.2, Chap. XIX.

The voltage e_b from plate to cathode, Fig. 2.2e, is the algebraic sum of the d-c supply voltage \bar{E}_{bb} , the modulation-frequency supply voltage $\hat{E}_l \sin \omega_l t$, and the instantaneous value of the modulated r-f voltage e_h developed across the tank circuit. The maximum instantaneous value of the plate-supply voltage e_{bb} is $\bar{E}_{bb} + \hat{E}_l$. From (2.7) the maximum amplitude of e_h is $k'(\bar{E}_{bb} + \hat{E}_l)$. In practice, k' is of the order of 0.9. During 100 per cent modulation, $\hat{E}_l = \bar{E}_{bb}$, and therefore the voltage from plate to cathode then reaches a maximum value $\bar{E}_{bb} + \hat{E}_l + k'(\bar{E}_{bb} + \hat{E}_l)$, which is of the order of 3.8 times \bar{E}_{bb} . This high voltage must be considered in choosing the plate-supply voltage \bar{E}_{bb} and in selecting capacitors and other elements to be used in the plate circuit.

Usually a network of some form is employed to transfer the r-f power developed in the plate circuit of the Class C amplifier to a load circuit. The emf induced in the secondary of the output transformer of Fig. 2.1 has essentially the same waveform as the tank-circuit current and is amplitude-modulated.

Under carrier conditions (no modulation), the power developed in the tank circuit is the carrier power and is

$$P_c = \frac{\bar{E}_h^2}{2R_L} = \frac{(k'\bar{E}_{bb})^2}{2R_L} \quad (2.8)$$

where R_L is the antiresonant resistance of the tank circuit at the fundamental, or carrier, frequency. During sinusoidal modulation, the power in a modulated wave increases by the factor $(1 + m_a^2/2)$, from (4.1), Chap. XIX. Then, when the Class C amplifier is sinusoidally modulated, the power delivered to the load or tank circuit increases to

$$P_L = \frac{(k'\bar{E}_{bb})^2}{2R_L} \left(1 + \frac{m_a^2}{2}\right) \quad (2.9)$$

assuming the impedance of the tank circuit to be the same at the side-band and the carrier frequencies.

It is desirable but not essential that the plate-supply current i_{bb} , Fig. 2.1, vary in proportion to e_{bb} . If this condition obtains,

$$\begin{aligned} i_{bb} &= k'' e_{bb} \\ &= k'' (\bar{E}_{bb} + \bar{E}_l \sin \omega t) \end{aligned} \quad (2.10)$$

where k'' is the constant of proportionality and the plate-supply voltage of (2.1) is assumed. Then with sinusoidal modulation the plate-supply current will consist of a d-c component $k'' \bar{E}_{bb}$ that is independent of the modulation and a modulating-frequency component whose amplitude $k'' \bar{E}_l$ is proportional to the amplitude of the modulating signal. The average power from the d-c source is equal to $\bar{E}_{bb} \bar{I}_{bb}$ and under the conditions of (2.10) is $k'' \bar{E}_{bb}^2$, independent of the modulation. The average power delivered by the modulating transformer T_2 , Fig. 2.1, is one-half the product of the amplitude of the modulating voltage \bar{E}_l and the amplitude $k'' \bar{E}_l$ of the modulation-frequency component of current. Therefore, under the conditions of (2.10), the average power supplied by the modulating transformer is $k'' \bar{E}_l^2/2$, or, from (2.4), $k'' m_a^2 \bar{E}_{bb}^2/2$, since $\bar{E}_l = m_a \bar{E}_{bb}$. The total average power supplied by the plate-supply source is

$$P_{bb} = k'' \bar{E}_{bb}^2 + k'' \frac{m_a^2}{2} \bar{E}_{bb}^2 \quad (2.11)$$

$$= k'' \bar{E}_{bb}^2 \left(1 + \frac{m_a^2}{2} \right) \quad (2.12)$$

The first term on the right of (2.11) is the d-c power, and the second term is the modulating power. Thus during sinusoidal modulation the source of the modulating signal must supply $m_a^2/2$ times as much power as the d-c source.

The plate-circuit efficiency is the ratio of power delivered to the load to power supplied by the plate supply. Hence, dividing (2.9) by (2.12),

$$\eta = \frac{P_L}{P_{bb}} = \frac{(k')^2}{2k'' R_L}$$

and the efficiency η is constant. This means that the power delivered to the load and the power dissipated at the plate increase by the same fractional amount when the modulation is increased, and both these increases are supplied from the modulating transformer since the average d-c input power remains constant.

For example, suppose that the plate input power of an unmodulated amplifier is 1,000 watts. If $\eta = 0.75$, the r-f power developed in the tank circuit is 750 watts. During sinusoidal modulation with $m_a = 1$, energy is delivered to the plate circuit by the modulation transformer at the average rate of 500 watts. Of this power, 375 watts becomes r-f side-band power. The total r-f output power is $750 + 375 = 1,125$ watts, an increase of 50 per cent over the unmodulated r-f power.

Under the conditions of (2.10), the modulating voltage \hat{E}_i of frequency f_i causes the plate-supply current to have a component of amplitude $k''\hat{E}_i$ in phase with \hat{E}_i . Therefore, the modulating transformer feeds a current into its load, proportional to and in phase with its secondary voltage. Thus the load on the modulating transformer simulates a resistor whose resistance is the ratio of voltage to current, or $\hat{E}_i/k''\hat{E}_i$. Since $\hat{E}_i = m_a\bar{E}_{bb}$, this ratio may be expressed as $\bar{E}_{bb}/k''\bar{E}_{bb}$, which is equal to $\bar{E}_{bb}/\bar{I}_{bb}$. Hence the d-c source and the modulation transformer both work into or have connected to them a load that is a resistance equal to $\bar{E}_{bb}/\bar{I}_{bb}$.

Side-band cutting (discussed in Sec. 10 of Chap. XIX) may be encountered if the Q of the plate tank circuit is too high, the oscillations in the tank circuit not dying away with sufficient rapidity when the plate-supply voltage decreases rapidly. These effects commonly are not encountered, for ordinarily the Q of the plate tank circuit is made sufficiently small by the resistance added to the circuit by the coupled load.

The advantages of plate modulation are high plate-circuit efficiency and ease of adjustment. A disadvantage is that a relatively large amount of low-frequency power is required, leading to heavy, bulky equipment.

When a screen-grid tube is plate-modulated, the screen voltage must be modulated to the same degree as the plate-supply voltage.

3. Grid Modulation of a Class C Amplifier.—Ideally, the amplitude of the current in the plate tank circuit of a Class C amplifier can be made to vary linearly with respect to the grid-bias voltage, the amplitude decreasing as the magnitude of the bias voltage is increased. Therefore an amplitude-modulated wave may be produced by applying a low-frequency modulating signal in series with the grid-bias supply.

For grid modulation the operating conditions are chosen so that ideally both the r-f plate tank current and the d-c plate-supply current vary linearly with the grid-bias voltage. Then, since the

average grid voltage is unchanged by the modulating voltage, the *average* plate current is unchanged, and the average power input from the d-c plate supply is unchanged by the modulation. Since the average r-f power output increases with modulation and since the average d-c power input remains constant, the plate dissipation decreases and the plate-circuit efficiency increases when the modulating signal is applied.

The operating conditions that result in linear modulation are such that, in the absence of modulation, the plate-circuit efficiency is of the order of 34 per cent. During 100 per cent modulation by a sinusoidal signal, the r-f output power increases by a factor of $\frac{1}{2}$, so that the plate-circuit efficiency increases to approximately 51 per cent. The carrier power, or the unmodulated output power, is about one-fourth as great as the carrier power that the same tube could deliver if plate modulation were employed.

The main disadvantages of grid modulation are low plate-circuit efficiency and the difficulty with which linear modulation is obtained. However, the equipment may be made light and compact since very little low-frequency modulating power is required.

4. Other Methods of Modulation of a Class C Amplifier.—If a pentode is employed, the modulation voltage may be injected into the suppressor-grid circuit. The basis of operation and the power relations are similar to those for control-grid modulation. However, adjustment of the amplifier is simpler. Fairly linear modulation up to 100 per cent may be obtained.

Modulation may be effected by injection of the modulating signal in series with the screen-supply voltage. This arrangement occasionally is used in low-power transmitters. Some power is required from the modulating source, and 100 per cent modulation usually cannot be attained with linearity.

Modulation voltage may be applied also in the cathode circuit of a Class C amplifier tube. The grid-bias and plate-supply voltages then are both altered during modulation. This is therefore a combination of plate-circuit and grid-circuit modulation. The adjustments required to secure linear modulation are of the same general nature as those required for grid-circuit modulation alone. The unmodulated, or carrier-component, output power obtained is greater than that which the same amplifier tube can deliver with ordinary grid-circuit modulation, while the amount of low-frequency power required is not as great as for plate-circuit modulation.

5. Amplitude-modulated Oscillator.—The output of an r-f oscillator may be amplitude-modulated by any of the methods used with Class *C* amplifiers, plate modulation being the most common method employed. Such modulation, however, usually is accompanied by unwanted changes in frequency because of changes that occur in the grid- and plate-polarizing potentials. The modulation characteristics (*i.e.*, curves relating the amplitude of the current in the plate tank circuit to the modulation voltage) are usually much less linear than those secured with a Class *C* amplifier. Modulated oscillators are used only where stability of frequency and low distortion are not essential or where more stable systems of modulation are not available.

6. Production of an Amplitude-modulated Wave by Use of a Nonlinear Impedance.—When a voltage that consists of one or

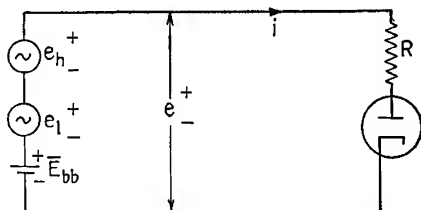


FIG. 6.1.—Circuit containing a nonlinear circuit element.

more frequencies is impressed on a circuit which contains a nonlinear impedance, the current produced in the circuit contains frequencies that are not present in the impressed voltage. This fact can be employed to advantage in modulation, in detection, and in frequency conversion.

For example, a circuit that contains a diode could be used to produce an amplitude-modulated wave. In the circuit of Fig. 6.1, a-c voltages e_h and e_l of frequencies f_h and f_l , having sinusoidal waveforms, and a steady direct voltage E_{bb} are applied in series across a diode and a resistor. The current-voltage characteristic of the circuit including the resistance is represented in Fig. 6.2. Under quiescent conditions the total applied voltage has the steady value E_{bb} , and the current has the steady value I_{bb} .

When the two a-c voltages are applied, the voltage e varies with time as shown in the lower part of Fig. 6.2 and is given by

$$e = E_{bb} + \hat{E}_h \sin \omega_h t + \hat{E}_l \sin \omega_l t \quad (6.1)$$

The corresponding values of current are shown at the right in Fig.

6.2. Because of the curvature of the current-voltage curve, the high-frequency variations are altered in amplitude during the modulation cycle.

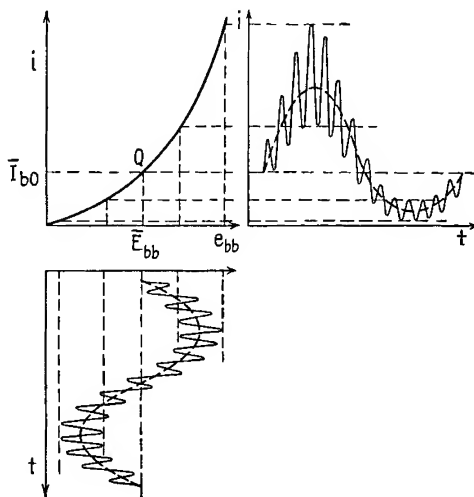


FIG. 6.2.—Operation as a modulator of a circuit containing a nonlinear circuit element.

Analysis of the current i is made as follows: Let e_0 and i_0 represent the deviations of the applied voltage and resulting current from their quiescent values. Thus, by this definition,

$$e_0 = e - \bar{E}_{bb} \quad (6.2)$$

and

$$i_0 = i - \bar{I}_{bb0} \quad (6.3)$$

From (6.1) and (6.2) the deviation of the voltage above and below the value \bar{E}_{bb} is

$$e_0 = \hat{E}_h \sin \omega_h t + \hat{E}_l \sin \omega_l t \quad (6.4)$$

The deviation of the current from the quiescent value in the vicinity of the Q point may be expressed by the Taylor's series

$$i_0 = ae_0 + \frac{b}{2!}e_0^2 + \frac{c}{3!}e_0^3 + \dots \quad (6.5)$$

in which the coefficients a , b , c , . . . are the first, second, third, . . . derivatives of the current with respect to voltage, evaluated at the Q point. The coefficients a and b represent the slope and the rate of change of the slope of the current-voltage curve at the Q point.

Substitution of (6.4) in (6.5) and performance of the indicated expansion show that terms in (6.5) of higher order than the second introduce components which constitute distortion and are not desirable. Therefore, for best results with this type of modulation, the current-voltage curve should be parabolic. Upon employing only the first and second terms of (6.5), it is found that

$$\begin{aligned}
 i_0 = & a\hat{E}_h \sin \omega_h t + a\hat{E}_l \sin \omega_l t \\
 & + \frac{b}{4} \hat{E}_h^2 - \frac{b}{4} \hat{E}_h^2 \cos 2\omega_h t \\
 & + \frac{b}{4} \hat{E}_l^2 - \frac{b}{4} \hat{E}_l^2 \cos 2\omega_l t \\
 & + \frac{b}{2} \hat{E}_h \hat{E}_l \cos (\omega_h - \omega_l)t - \frac{b}{2} \hat{E}_h \hat{E}_l \cos (\omega_h + \omega_l)t \quad (6.6)
 \end{aligned}$$

In (6.6) there are terms having the angular frequencies ω_h , $\omega_h - \omega_l$, $\omega_h + \omega_l$, which constitute the carrier and side frequencies of an amplitude-modulated wave. These terms are

$$i = a\hat{E}_h \left[\sin \omega_h t + \frac{b}{2a} \hat{E}_l \cos (\omega_h - \omega_l)t - \frac{b}{2a} \hat{E}_l \cos (\omega_h + \omega_l)t \right] \quad (6.7)$$

A selective amplifier can be driven by the voltage developed across R in order to separate the carrier and side-band terms from the other terms in (6.6). By this means, an amplitude-modulated current can be produced having the form

$$i = \hat{I}_h (1 + m_a \sin \omega_l t) \sin \omega_h t \quad (6.8)$$

where the degree of modulation m_a is equal to $\frac{b}{a} \hat{E}_l$. A similar result can be obtained by using a circuit tuned to the frequency f_h in place of R .

7. Triode Square-law Modulator.—A Class A amplifier, operated in the negative-grid region where the i_b - e_c characteristic curves are nonlinear may be used to produce an amplitude-modulated wave. Such a device is known as a square-law or van der Bijl modulator. The basic circuit is shown in Fig. 7.1. Capacitor C_1 tunes the r-f input circuit to the frequency f_h ; capacitor C_2 by-passes the r-f currents around the modulation transformer T_2 ; and capacitor C_3 tunes the output circuit to the frequency f_h . The output circuit has a resonance curve that is broad enough to include the frequencies $f_h + f_l$ and $f_h - f_l$. The grid-bias voltage \bar{E}_{cc} is

of such a value as to place the Q point in the square-law region of the tube characteristics.

During operation, the grid is always negative, but its potential varies in the manner of e_{bb} in Fig. 6.2. The corresponding values of plate current are similar to those in Fig. 6.2. Although the amplitude of the r-f swing in grid potential remains constant during

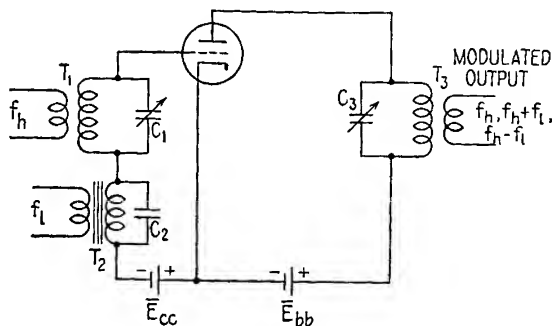


FIG. 7.1.—Single-tube square-law modulator circuit.

modulation, the amplitude of the r-f component in the plate current is increased when the low-frequency voltage across T_2 , Fig. 7.1, opposes \bar{E}_{cc} and is decreased when it aids \bar{E}_{cc} .

8. Balanced Modulator.—An arrangement known as a *balanced* modulator is shown in Fig. 8.1. It consists of two single-tube modulators connected back to back and has properties of consider-

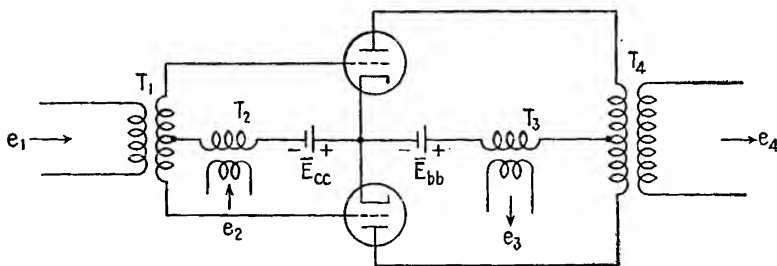


FIG. 8.1.—Balanced modulator circuit.

able practical value. The two input voltages e_1 and e_2 may have any frequencies. The input voltage e_2 is applied to the two grids in the same phase. The input voltage e_1 is applied to the upper grid 180° out of phase with the input voltage e_1 as applied to the lower grid. Output voltage e_3 arises from changes in the two plate currents flowing in the same direction through the primary of transformer T_3 . Output voltage e_4 arises from changes in the plate cur-

rents flowing in opposite directions in the primary of T_4 . The voltages e_3 and e_4 therefore differ from each other, and either may be used as the output voltage.

By assuming input voltages of sinusoidal waveform and square-law operation of the tubes as modulators, the components of the output voltages e_3 and e_4 may be determined by an analysis similar to that used in Sec. 6. The component frequencies found in the two outputs under various input conditions are tabulated in Table 8.1. The first row in Table 8.1 applies when the circuit is operated like a push-pull amplifier, one signal being applied at input 1. Even though each half of the circuit is nonlinear, the square-law-distortion output components do not appear at output 4. Row 2 represents the operation of two amplifier tubes in parallel.

The third row represents the fact that application of a high-frequency voltage at input 1 and a low-frequency modulating voltage at input 2 results in an amplitude-modulated wave at output 4 with no extraneous frequencies. This represents a distinct improvement over the output of a single-tube square-law modulator.

TABLE 8.1.—INPUT AND OUTPUT FREQUENCIES OF A BALANCED SQUARE-LAW MODULATOR CIRCUIT

Row	e_1	e_2	e_3	e_4	Remarks
1	ω_i	0	$2\omega_i$	ω_i	Push-pull amplifier
2	0	ω_i	$\omega_p, 2\omega_i$	0	Two tubes in parallel
3	ω_h	ω_i	$\omega_p, 2\omega_p, 2\omega_h$	$\omega_h, \omega_h + \omega_i, \omega_h - \omega_i$	e_4 is an amplitude-modulated wave
4	ω_i	ω_h	$\omega_h, 2\omega_p, 2\omega_h$	$\omega_p, \omega_h + \omega_i, \omega_h - \omega_i$	Carrier suppressed
5	ω_h, ω_i	0	$2\omega_h, 2\omega_i, \omega_h + \omega_p, \omega_h - \omega_i$	ω_h, ω_i	Carrier suppressed

The fourth and fifth rows represent methods by which r-f side bands minus the carrier-frequency component may be created. These properties of the balanced square-law modulator make it very useful in applications such as suppressed-carrier and single-side-band transmission, speech scrambling, and frequency modulation by the Armstrong method.

II. METHODS OF ANGLE MODULATION

9. General.—An angle-modulated wave may be produced by altering the frequency of an oscillator or by means of a phase-shifting network. Transmitters whose output is a phase-modulated wave are seldom used since a phase-modulated wave (defined in

Sec. 6 of Chap. XIX) does not utilize a given frequency channel as effectively as a frequency-modulated wave, as explained in Sec. 8 Chap. XIX.

The frequency of an r-f oscillator may be varied by means of a "condenser microphone," Fig. 9.1. The capacitive reactance of the tuned circuit is altered in accordance with sound waves entering the microphone. The frequency of oscillation varies inversely as the square root of the capacitance of the circuit; but the variations in capacitance are extremely small, and the variations in frequency are practically proportional to the successive values of the sound wave. This method is mechanically and electrically impractical, but it illustrates the fundamental principle of the frequency-modulated oscillator.

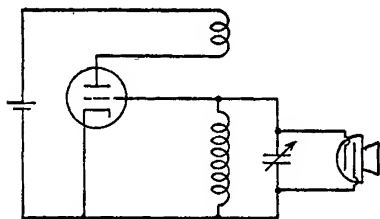


FIG. 9.1.—Condenser-microphone method of frequency modulation.

The frequency of an oscillator may be controlled by a variable reactance produced by a reactance-tube circuit connected across the tank circuit of the oscillator. The circuit, described in Sec. 10, simulates a capacitance or an inductance whose magnitude is controlled by a

low-frequency signal voltage applied to the reactance tube. In many frequency-modulated transmitters, this arrangement is used to control the frequency of an oscillator operating at a center frequency considerably below that of the output carrier; the frequency-modulated-oscillator output is passed through frequency-multiplying stages, in which the center frequency and the frequency deviations are multiplied by the same factor. With this method of producing frequency modulation the value of the center frequency is subject to many influences in the oscillator and reactance-tube stages (such as variations in polarizing or supply voltages, mechanical movement of parts, temperature changes) that lead to drift or change of the center frequency. Center-frequency stability approaching that of crystal-controlled transmission may be obtained, however, by use of a stabilization system.

Modulation may be accomplished also by use of a phase-shift circuit rather than a reactance tube. The steady r-f output of a crystal-controlled oscillator operating at a frequency considerably below that of the output carrier is applied to a phase-shift circuit. The phase of the r-f output of the phase-shift circuit relative to its

r-f input is altered by means of and in direct proportion to low-frequency voltage applied to the phase-shift circuit. The output of the phase-shift circuit is sent through frequency-multiplying stages before transmission. If the low-frequency wave representing the intelligence were applied directly to the phase-shift circuit, the output of the transmitter would be a phase-modulated wave. However, in these transmitters, the low-frequency wave is first passed through an integrator circuit as explained in Sec. 9 of Chap. XIX. The output of the integrator, or "distorter," circuit has a waveform identical with that of the phase shifts which would accompany a frequency-modulated wave created by applying the original unintegrated wave to a reactance tube controlling the frequency of an oscillator. When the integrated wave is applied to the phase-shift circuit, the phase shifts it creates are accompanied by frequency shifts that are proportional to the instantaneous value of the modulating wave. Hence the output of the transmitter is frequency-modulated in accordance with the unintegrated modulating wave. The integrator circuit has a $1/f_i$ frequency-response characteristic. The integrator may be an amplifier whose voltage amplification is inversely proportional to the audio frequency. It may be an attenuator that attenuates the audio signal in proportion to the audio frequency.

The primary advantage of the frequency-modulation systems that employ a phase-shift circuit and integration of the modulating wave lies in the fact that a separate control system for stabilization of the center frequency is unnecessary since a very stable oscillator, such as a crystal oscillator, may be employed as the source of the center frequency. The primary disadvantage is that, in the simpler transmitters of this type, the frequency shifts at the output are relatively smaller.

The higher frequency components of the modulating wave ordinarily are weak compared with the lower frequency components. In frequency-modulation broadcast transmission, the amplitudes of the higher frequency audio components usually are raised, or emphasized, by means of an emphasis, or accentuation, circuit. They may be raised considerably without much danger of the frequency swings they create exceeding the legal restrictions. The higher components are deemphasized after the signal is detected in the receiver. Noise voltages are thus attenuated, and the signal-to-noise ratio of the system is improved. In the 42- to 50-mc frequency-modulation band the standard emphasis (or preemphasis,

as it is commonly called) is in accordance with the impedance-frequency characteristic of a series inductance-resistance network having a time constant of 100 microsec, Fig. 9.2*a*. In the 88- to 108-mcps band the standard corresponds to a time constant of 75 microsec, Fig. 9.2*b*. The deemphasis in the receiver corresponds to the impedance-frequency characteristic of a parallel resistance-capacitance network having the same time constant as that governing the emphasis in the transmitter.

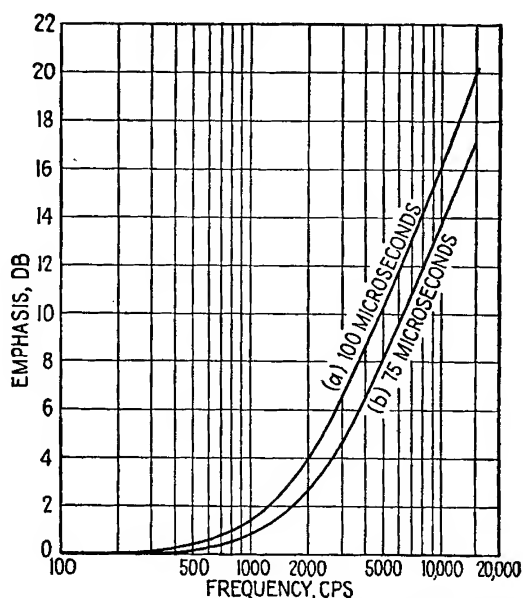


FIG. 9.2.—Emphasis curves corresponding to the impedance-frequency characteristics of series LR circuits having time constants of 100 and 75 microsec.

10. Reactance-tube Modulation of an Oscillator.—A reactance-tube circuit is shown in Fig. 10.1, attached at the points AB to a Hartley oscillator. The reactance-tube circuit as a whole, *i.e.*, everything to the right of the points AB , simulates (in this case) a small capacitance C_{AB} in parallel with C . The approximate frequency of oscillation is therefore

$$f \doteq \frac{1}{2\pi \sqrt{L(C + C_{AB})}} \quad (10.1)$$

The value of C_{AB} may be altered by the application of low-frequency voltage at any one of several points in the reactance-tube circuit.

How the tube T_2 in Fig. 10.1 simulates a reactance will now be explained.

The r-f voltage E across the tank circuit causes a very small r-f current i_1 in the voltage divider C_1R_1 . The reactance of C_1 at the frequency of oscillation is of the order of

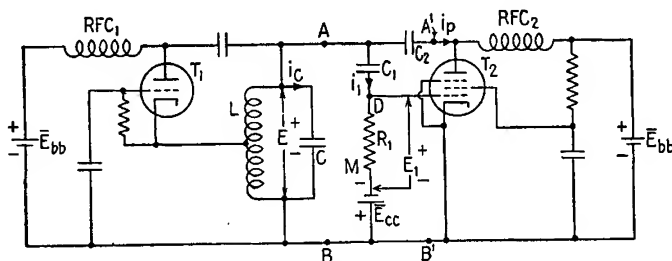


FIG. 10.1.—Reactance-tube circuit attached to a Hartley oscillator.

ten times the resistance of R_1 . Therefore E_1 , the r-f potential difference developed across R_1 , leads E by almost 90° and is approximately

$$E_1 \doteq jR_1\omega C_1E \quad (10.2)$$

The grid potential of the reactance tube is the sum of E_1 and the bias voltage \bar{E}_{cc} . The bias voltage ensures Class A_1 operation.

The r-f grid voltage E_1 , leading E by 90° causes the plate current of the reactance tube to have an r-f component that also leads E by 90° . This current is prevented from flowing through the plate-supply system by the choke coil RFC_2 ; it is in phase with the current i_c of the capacitor C in the tank circuit, and the tube therefore simulates a capacitor in parallel with C . Increasing or decreasing the quadrature component of the reactance-tube plate current produces the same effect as increasing or decreasing C . The magnitude of the quadrature component may be altered by changing the g_m of the tube, for example by injecting a modulating signal voltage at M , Fig. 10.1, in series with \bar{E}_{cc} .

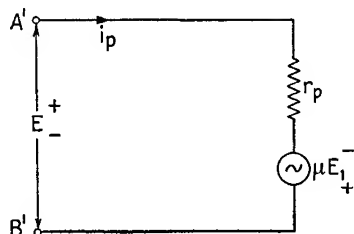


FIG. 10.2.—Equivalent plate circuit of a reactance tube.

The r-f equivalent plate circuit of the reactance tube at the points $A'B'$ is shown in Fig. 10.2 and is a series circuit in which

the voltage E and the equivalent voltage μE_1 are in series with r_p . Then the r-f part of the plate current is

$$I_p = \frac{E + \mu E_1}{r_p} \quad (10.3)$$

From (10.2),

$$I_p \doteq \frac{E + j\mu R_1 \omega C_1 E}{r_p} \quad (10.4)$$

The inphase current E/r_p represents only a slight dissipative load on the tank circuit since the r_p of a pentode is high. The second term represents a reactive current that is variable, since the amount of this leading current can be changed by varying g_m .

In the circuit of Fig. 10.1 the total leading current at the terminals AB is approximately

$$j\omega C_1 E + jg_m R_1 \omega C_1 E$$

which may be equated to $j\omega C_{AB} E$. Then

$$j\omega C_{AB} E \doteq j\omega C_1 E + jg_m R_1 \omega C_1 E \quad (10.5)$$

so that

$$C_{AB} \doteq C_1 + g_m R_1 C_1 \quad (10.6)$$

$$\doteq C_1 + C_{RT} \quad (10.7)$$

where C_{RT} is the effective capacitance of the reactance tube only. Even though approximations have been employed, it is evident from (10.6) that the value of C_{AB} is linearly related to the transconductance g_m . Over a limited range, g_m varies linearly with respect to the grid-bias voltage. Therefore, the value of C_{AB} is altered linearly with respect to a low-frequency signal voltage injected at a point such as M . Since the frequency f usually is changed by less than 0.2 per cent, the frequency deviations are practically proportional to the successive values of the modulating signal.

A primary feature of a reactance-tube circuit is the network used to obtain a fraction of the applied r-f voltage that is shifted in phase approximately 90° . Four such networks are shown in Fig. 10.3. Analyses of these circuits may be found elsewhere.¹ Two of these networks cause the reactance tube to simulate a capacitance, and two an inductance. An inductance is simulated when the r-f grid voltage lags rather than leads the tank-circuit voltage. When the reactance-tube circuit is inductive, the approximate frequency

¹ A. HUND, "Frequency Modulation," pp. 155-174, McGraw-Hill Book Company, Inc., 1942.

of oscillation is

$$f \doteq \frac{1}{2\pi \sqrt{\left(\frac{L_{AB}L}{L_{AB} + L}\right)C}} \quad (10.8)$$

where L_{AB} is the effective shunt inductance due to the reactance-tube circuit. Increasing the g_m of the tube decreases L_{AB} and increases the frequency of oscillation.

Low-frequency modulation voltage may also be applied to the screen grid or to the suppressor grid. Some multigrid tubes have two control grids, to one of which the modulating signal may be applied.

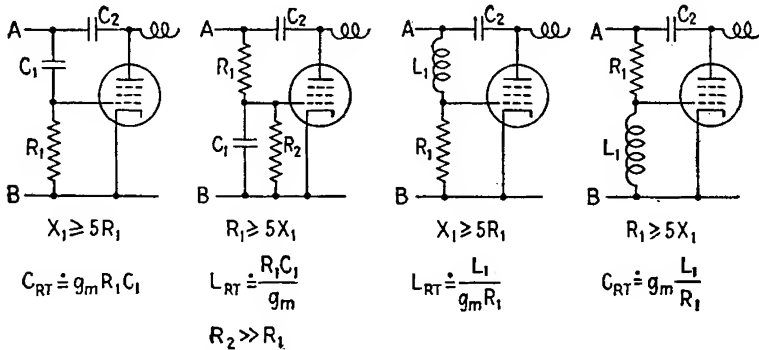


FIG. 10.3.—Reactance-tube circuits. C_{RT} , or L_{RT} , is the effective input capacitance, or inductance, neglecting the shunting effect of the voltage-divider network.

A block diagram of a transmitter that employs a reactance tube is shown in Fig. 10.4. The frequencies given in this description are representative values for pre-war commercial frequency-modulation broadcasting. The modulation voltage is passed through an emphasis or accentuator circuit and then into the reactance-tube circuit. The oscillator operates at, say, 5 mcps. The maximum frequency deviation at the oscillator, produced by the loudest modulating-voltage input to the transmitter, may be 8,000 cps. The frequency-modulated output of the oscillator then is passed through frequency-multiplier stages, in this case two triplers. The center frequency f_h at the output is 45 mcps, which is 3 times 3, or 9 times the center frequency of the oscillator. The frequency deviations are magnified in the same ratio. For the signal of maximum strength, the value of Δf at the antenna is 9 times 8, or 72 keps. The modulation index m_f then has the value $8,000/f_i$ at the oscillator and $72,000/f_i$ at the antenna.

To prevent drift of the center frequency, a "sample" of the output of the transmitter is mixed with the 42-mcps output of a crystal-controlled oscillator and frequency-multiplier system. The output of the mixer (frequency converter) is a frequency-modulated wave having a center frequency of 3 mcps (if the transmitter is operating at its correct center frequency) and a maximum frequency deviation of 72 keps (for the maximum signal assumed). The mixer is followed by a discriminator¹ tuned to 3 mcps. When the transmitter is operating on its correct center frequency, the output of the discriminator is an alternating voltage that is a reproduction of the modulating signal. If the center frequency of the trans-

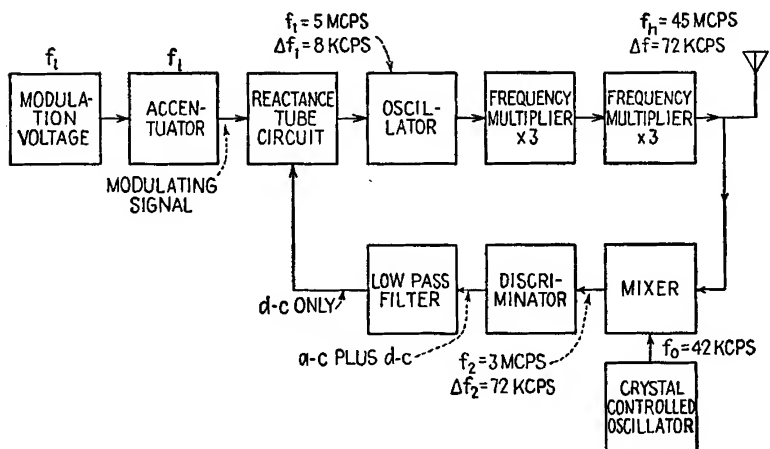


FIG. 10.4.—Block diagram of a frequency-modulation transmitter employing a reactance tube as a modulating device.

mitter should drift upward, the center frequency of the mixer output becomes greater than 3 mcps, and the output of the discriminator has a d-c component in addition to the modulating-frequency components. A low-pass filter following the discriminator passes only the d-c component, which is applied to the grid of the reactance tube in such a direction as to decrease the center frequency of the oscillator to the correct value. If the center frequency of the transmitter should drift down instead of up, a frequency-correction voltage of opposite polarity is developed.

11. Production of Frequency Modulation by Phase Modulation.

A relatively simple frequency-modulation transmitter of the phase-shift variety is shown in block-diagram form in Fig. 11.1.

¹ See Chap. XXI, Sec. 22, and Chap. XXIII, Sec. 12.

A crystal-controlled oscillator drives an LC circuit tuned to the crystal frequency. The LC circuit is coupled to the oscillator by means of a phase-shift network, which is essentially a phase-splitting circuit; its output voltage leads (or lags) its input voltage by a certain number of degrees. Variation in the potential applied to one or more vacuum tubes in the phase-shift circuit changes, at a

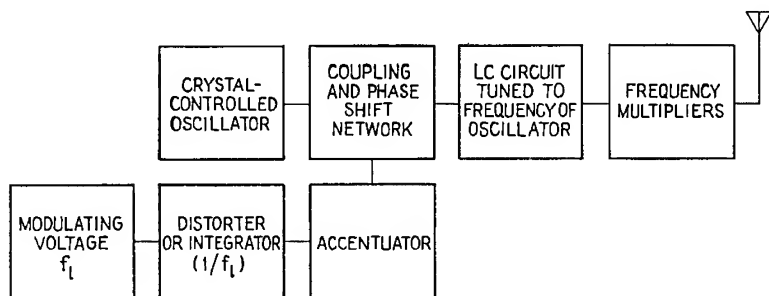


FIG. 11.1.—Block diagram of a frequency-modulation transmitter employing a phase-shift modulator.

low-frequency rate, the value of the angle of lead or lag between the input and output voltages of the coupling network. Since the oscillator is rigidly controlled by the crystal, the voltage across the LC circuit is advanced and retarded in phase at the low-frequency rate. In addition to the phase shifts produced, small changes in

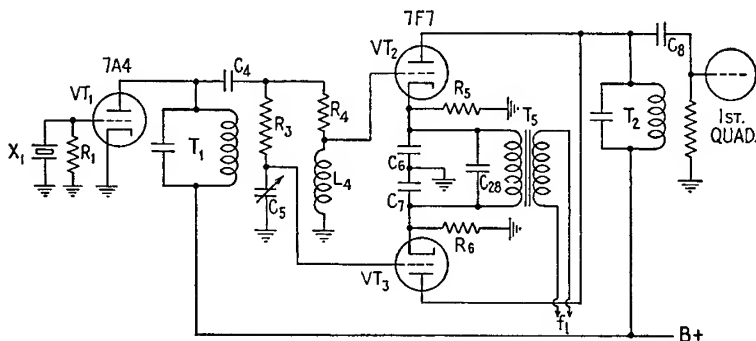


FIG. 11.2.—An example of a phase-shift modulator circuit.

the amplitude of the voltage across the LC circuit may also occur but are removed by limiting action in the subsequent stages.

The phase-shift circuit in the type of transmitter described may have many forms. The circuit¹ of Fig. 11.2, designed for

¹ A. H. QUIST, JR., Single-unit Mobile FM Equipment, *FM Mag.*, March and April, 1942.

police communication, has very few parts but illustrates the principles common to many phase-shift modulators. VT_1 is a triode employed in a crystal oscillator whose frequency of oscillation is the output frequency divided by 32. The r-f voltage developed across the plate tank circuit T_1 is also applied to the voltage-divider circuits R_3C_5 and R_4L_4 . With the resistance equal to the reactance in these voltage-divider circuits, the r-f voltage applied to the grid of VT_2 leads the r-f plate voltage of VT_1 by approximately 45° , and the r-f voltage applied to the grid of VT_3 lags the r-f plate voltage of VT_1 by a similar amount. In the absence of modulation these voltages are equal, and the vector sum of the r-f plate currents of VT_2 and VT_3 is in phase with the r-f plate voltage of VT_1 . This is

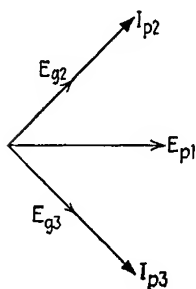


FIG. 11.3.—Radio-frequency components of the grid voltages and plate currents of tubes VT_2 and VT_3 in Fig. 11.2.

illustrated in Fig. 11.3, where I_{p2} and I_{p3} , the r-f plate currents of VT_2 and VT_3 , are assumed to be in phase with the r-f voltages applied to the grids. The modulating signal is applied to the cathodes of VT_2 and VT_3 , so that, by cathode modulation, I_{p2} and I_{p3} may be varied. When the modulating signal drives the cathode of VT_2 negative with respect to ground potential, I_{p2} increases. Simultaneously, the cathode of VT_3 is driven positive with respect to ground potential, and I_{p3} decreases. The resultant current then leads E_{p1} of Fig. 11.3. When the modulating signal reverses, the changes in plate current also reverse, so that the resultant of I_{p2} and I_{p3} lags E_{p1} . Thus the r-f current in the tank circuit T_2 , Fig. 11.2, is shifted in phase by an amount controlled by the modulating signal appearing across the secondary of the transformer T_5 .

The characteristics of the audio-frequency system including transformer T_5 and its audio-frequency load are designed to supply the $1/f_i$ characteristic necessary with this type of modulator. Some accentuation is also incorporated into this circuit. The requirements as to fidelity are not very severe in police communication; this transmitter and its associated receiver are designed for the audio-frequency band from 500 to 3,000 cps.

The modulator stage is followed by two quadruplers, one doubler, and a final power-amplifier stage of conventional construction. These are not shown in Fig. 11.2 but are indicated functionally in the block diagram of Fig. 11.1.

This transmitter is designed for frequency swings of approximately ± 15 keps at modulation peaks. Suppose, then, that the output carrier frequency is 32 mcps and that the frequency deviation with a 1,000-cps sinusoidal test signal is ± 12 keps, giving the spectrum indicated in Fig. 8.2a, Chap. XIX. Since the frequency multiplication is 32, the center frequency and the frequency swings at the output of the modulator are $\frac{1}{32}$ of the output values, or 1 mcps and ± 375 cps, respectively. Then at the output of the modulator,

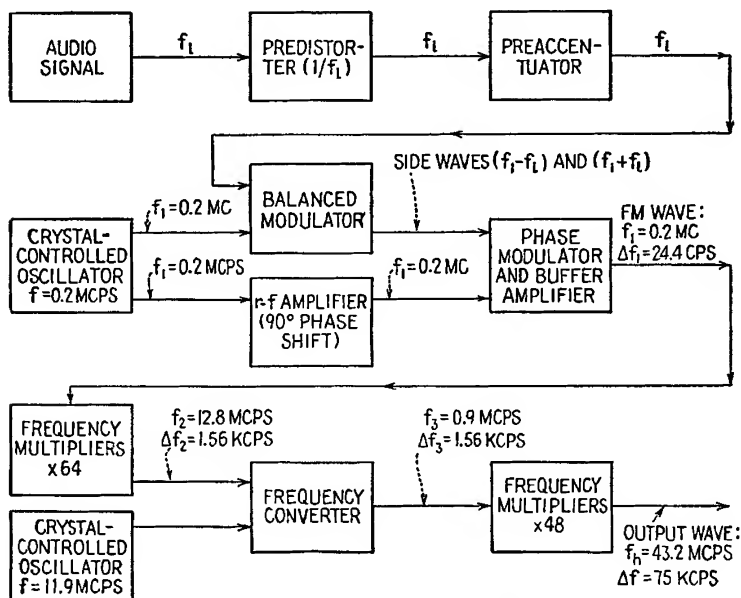


FIG. 11.4.—Simplified block diagram of the original Armstrong broadcast frequency-modulation transmitter. Values of Δf are for the strongest permissible modulating signal.

$m_f = \Delta f/f_i = 0.375$, so that the maximum departure in phase is 0.375 radian.

The special advantages of frequency modulation are increased by the use of wide frequency swings. The early Armstrong¹ wide-swing transmitter for broadcast service is shown in block-diagram form in Fig. 11.4. The audio-frequency modulating wave is passed through the distorting, or integrating, circuit, where the $1/f_i$ characteristic is imparted, then through the accentuator circuit, and then applied as the modulating signal to a balanced modulator operating on the basis of row 4 of Table 8.1. The output voltage is passed

¹ E. H. ARMSTRONG, *Proc. I.R.E.*, **24**, 689, May, 1936.

through a filter network to eliminate the low-frequency modulating signal. The side bands resulting from the amplitude modulation in the balanced modulator are recombined with the carrier after the carrier has been shifted 90° in phase. The result is a phase-modulated wave having the desired frequency swing.

The operation of this modulator for a sinusoidal modulating signal is as follows: The balanced modulator produces the two side frequencies characteristic of amplitude modulation. The carrier and side-frequency vectors for amplitude modulation have the relative positions indicated in Fig. 5.1, Chap. XIX. If the carrier is shifted in phase by 90° , the relative positions of the vectors are as indicated in Fig. 11.5a. The resultant, or sum, is a vector that shifts from

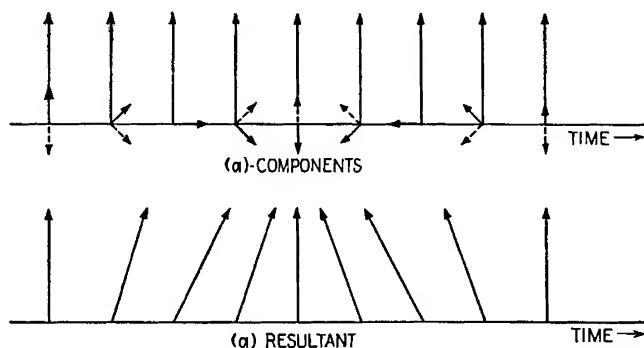


FIG. 11.5.—Component vectors and their resultant after the carrier vector has been advanced 90° .

one side to the other of the center position. If the angle of this shift is small, the change in angle can be made proportional to the value of the modulating signal. The distortion is less than 5 per cent for phase shifts not exceeding 25° . A 90° phase shift in the carrier-frequency voltage may be secured by an untuned r-f transformer having a large resistance in series with the primary coil. Some amplitude modulation is produced in this phase modulator, but it is not very great and is reduced by succeeding Class C stages, in which the output is fairly independent of the input signal when the input signal is sufficiently large.

In order to develop a large frequency deviation the output of the modulator is frequency-multiplied, converted to a lower center frequency, and multiplied further. The prewar frequency-modulation broadcast band is from 43 to 50 mcps. The legal limit to the frequency deviation, or maximum frequency swing, is 75 kcps. In

Fig. 11.4 the center frequency and the frequency deviations are marked for an output carrier frequency of 43.2 mcps with a frequency deviation of 75 keps. At the output of the phase-modulator buffer amplifier, in which the vector addition indicated in Fig. 11.5 is performed, the center frequency is 200,000 cps, and the frequency deviation is 24.4 cps. At a modulating frequency of 100 cps the value of m_f at the phase modulator is 0.244 so that the maximum deviation in phase is 0.244 radian, or 14° . At a modulating frequency of 15,000 cps the corresponding deviation is only 0.00163 radian. At the output of the transmitter the maximum phase deviation is 750 radians for a modulating frequency of 100 cps and 5 radians for a modulating frequency of 15,000 cps. The large frequency multiplication employed in this transmitter requires very careful control over the frequency of the crystal-controlled oscillators in order to stabilize the center frequency at the output of the transmitter.

CHAPTER XXI

DETECTION

1. Introduction.—The purpose of a radio receiver is to respond to the electrical forces due to currents in a transmitting antenna in such a manner as to convey desired information. The information may be a series of dots and dashes, as in telegraphy, or an audio- or video-frequency signal, as in voice or picture transmission. The purpose of the detector circuit in the radio receiver is to respond to the impressed radio-frequency voltage in such a way that a current is generated whose variations, owing to modulation at the transmitter, produce some perceptible effect.

The American Standard Definitions of Electrical Terms defines¹ detection as follows: "Detection is any process of operation on a modulated signal wave whereby the signal imparted to it in the modulation process is obtained." The word "demodulation" is also descriptive of the work performed by a detector circuit. The American Standard Definitions contains the following definition: "Demodulation is the process whereby a wave resulting from modulation is so operated upon that a wave is obtained having substantially the characteristics of the original modulating wave." For audio- or video-frequency work, detection and demodulation are synonymous. Detection is the broader term and will be used in this chapter.

In modern practice, detection is a process of rectification, producing a low-frequency current that is used to create acoustic or visual effects. Any device capable of rectifying a radio-frequency voltage can be used in a detector circuit. Such devices are non-linear, because the instantaneous current through them is not proportional to the instantaneous voltage impressed upon them. In high-frequency radio, vacuum tubes and crystals are employed; for low radio frequencies such as are used for carrier-current telephony, copper oxide rectifiers also are employed. Some rectifying devices cannot be used in high-frequency radio applications because

¹ American Standard Definitions of Electrical Terms, ASA C42-1941, American Institute of Electrical Engineers, 1942.

their large dimensions cause them to have too much shunting capacitance. When triodes or pentodes are used, the rectification may take place in the grid circuit, the detection being called grid-circuit detection, or in the plate circuit, the detection being called plate-circuit detection.

2. Square-law and Linear Detection.—Since a rectifying device is nonlinear, its current-voltage characteristic must be curved or kinked. In the diode, the characteristic may be represented to a close approximation by assuming that for low plate voltages the i_b - e_b characteristic is parabolic. If the voltage impressed on the diode is small, the rectification is such that the increase in the average current is proportional to the square of the amplitude of the r-f voltage. This is a characteristic of square-law detection and is the type of detection usually encountered when the r-f voltage is small.

On the other hand, if the impressed r-f voltage is large, this causes the plate to become so negative with respect to the cathode that there is no plate current during a large part of the r-f cycle. In a correctly designed circuit, the variations of the average current are proportional to the variations of amplitude of the r-f voltage. This is called linear detection and is the type of detection usually encountered when the r-f voltage is large. For voltages in excess of 1 or 2 volts, the detection is usually linear if the rectifier has its cutoff in the region of zero voltage.

LARGE-SIGNAL (LINEAR) DIODE DETECTION

3. Large-signal Rectification.—At present, the diode is the most important rectifying device used in large-signal detector circuits. As a linear detector, the diode is employed as a rectifier to prevent the flow of current when the plate potential is negative with respect to the cathode. Actually, there is a slight plate current in a thermionic diode at zero and negative plate potentials, due to initial velocity of electron emission and differences in contact potentials. For example, a microammeter connected between plate and cathode of one 6H6 diode indicates a current of the order of 100 microamperes, there being no external batteries or sources of emf in the circuit. This current at zero plate voltage varies widely in different tubes, owing possibly to differences in contact potentials. The direction of this current is the same as when the plate voltage is positive. However, for plate potentials more negative than 1 or 2 volts, the current is negligible. At large negative plate voltages

there is no reverse current except for a minute amount due to the ionization of the residual gas within the tube. In any case, the rectifier in a linear-detector circuit functions best when the reverse current is zero and when the rectifier, in response to an impressed a-c voltage, begins and ceases to conduct at the instants when the voltage across its terminals is zero. The diode will be assumed to possess this property when used as a linear detector.

The object of linear detection is to produce a current or voltage whose average value measured from the quiescent value is proportional to the amplitude of the applied r-f voltage. If the applied voltage is amplitude-modulated, the object is to develop a voltage whose value fluctuates in the same manner as the amplitude of the modulated wave.

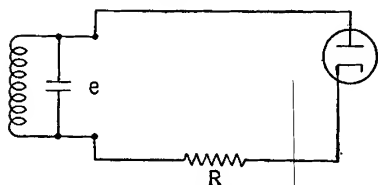


FIG. 3.1.—Diode with pure resistance load.

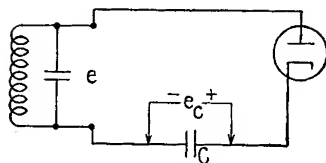


FIG. 3.2.—Diode with pure capacitance load.

In certain respects a rectifier in series with a pure resistance would accomplish these objects. Consider the circuit of Fig. 3.1, in which for this discussion the resistance R is assumed large compared with the resistance of the diode when conducting. Half-wave rectification takes place; in response to an unmodulated or a modulated voltage the current follows the instantaneous r-f voltage during its positive values. The average current over one r-f cycle is $1/\pi$ times the peak value of the rectified half waves.¹ With R large, and with an unmodulated voltage impressed, the d-c voltage across R is $1/\pi$ times the amplitude of the impressed voltage. If the amplitude of the impressed voltage is increased and decreased by modulation, the d-c voltage across R will increase and decrease similarly, yielding the desired audio- or video-frequency voltage.

This form of detection apparently is ideal but actually is not. If the rectification is half-wave, the instantaneous peak voltage across R is equal to the amplitude of the applied voltage, so that the r-f voltage across R is greater than the d-c or audio voltage. This r-f voltage would have to be filtered out to prevent its disturbing the following amplifier tubes in the receiver.

¹ Chap. IX, Fig. 10.9 and Eq. (10.9).

In practice, the resistor R is shunted by capacitance due to the detector circuit itself together with the input capacitance of the devices into which the output of the detector is fed. If the r-f reactance of this shunting capacitance is appreciable compared with R , the action of the circuit is completely changed. How complete this change is may be shown most dramatically by going to the opposite extreme and considering the diode load to be a pure capacitance, Fig. 3.2. Let an r-f signal voltage e be applied to this circuit. The plate of the diode is driven positive with respect to the cathode during the first positive half cycle of e , and the diode conducts. The flow of electrons will charge the capacitor C with a polarity such that the plate acquires a negative bias voltage with

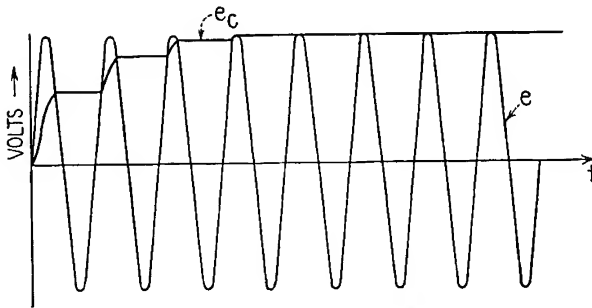


FIG. 3.3.—Comparison of e_C with the applied voltage e in the circuit of Fig. 3.2.

respect to the cathode, and during the next positive half cycle of applied voltage the diode will not conduct until the positive instantaneous value of the applied voltage exceeds the negative bias voltage previously developed across C . During the time when the diode does conduct, the capacitance C accumulates additional charge and biases the plate still more negatively. This action will continue until the magnitude of the voltage to which the capacitor is charged is equal to the amplitude of the r-f voltage. Figure 3.3 illustrates how the magnitude of the voltage across C increases to the amplitude, or peak, value of the applied r-f voltage. After the capacitor C is charged to the amplitude, or peak, value of the applied voltage, the plate can no longer be driven positive and the capacitor C can be charged no further. This is indicated in Fig. 3.4, where the variations in e_b are plotted. If the applied voltage were removed, the charge on C would remain.

Note that the capacitor becomes charged to the peak value of the applied voltage because there is no reverse current in the diode,

and not because of any particular form of the diode i_b - e_b characteristic during conduction. This is of importance in some linear-detection circuits and vacuum-tube voltmeter circuits, in that the peak-voltage indication does not require a linear i_b - e_b characteristic during conduction.

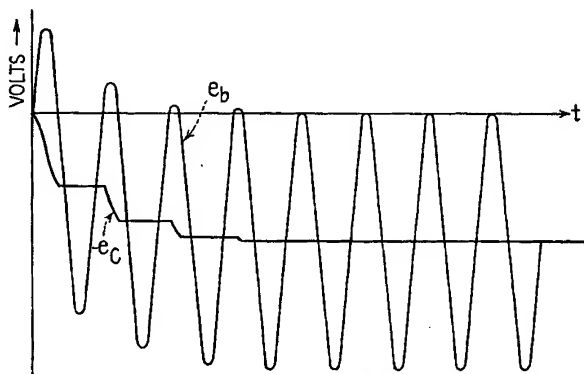


FIG. 3.4.—Variations in e_b corresponding to conditions of Fig. 3.3.

4. Diode with RC Load.—If resistance is in parallel with the capacitance C of Fig. 3.2, the result is the extremely useful circuit of Fig. 4.1. The capacitance C is usually large enough to by-pass R at radio frequencies so that practically all the impressed r-f voltage appears across the diode. Even with R present, the tendency of the circuit to charge C to the peak value of the applied voltage persists. The capacitance will be charged to the peak value if

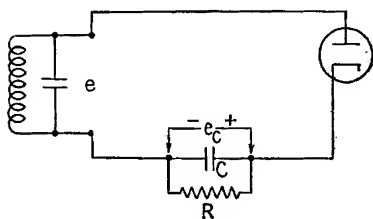


FIG. 4.1.—Diode with RC load.

the resistance of the diode during the conducting period is small compared with R . Owing to the time constant of the RC circuit, the capacitance tends to hold this charge, so that, when the instantaneous r-f voltage falls below its peak value, the negative voltage built up across C causes the diode plate to become negative with

respect to the cathode and conduction ceases. This action is the same as that discussed in Chap. XVII on Rectifiers and Power Supplies, except that here the voltages are lower and the frequencies much higher than those commonly encountered in power filters and the resistance of the rectifier is such that it cannot be neglected.

After the diode has ceased to conduct, the capacitance C receives no more charge and discharges through R until the next positive half cycle of the r-f voltage again drives the plate of the diode sufficiently positive to make the instantaneous (charging) current through the diode greater than the (discharging) current through R . This action is illustrated in Fig. 4.2, where the magnitudes of e_c and \bar{E}_C are compared with the amplitude \hat{E} of a sinusoidal applied voltage e . If the resistance e_b/i_b of the diode when conducting is small compared with R and if the time constant of the RC load is large compared with the period of the r-f cycle, the average voltage \bar{E}_C across the capacitance almost equals the amplitude \hat{E} of the applied voltage. If these conditions do not hold, \bar{E}_C may be much

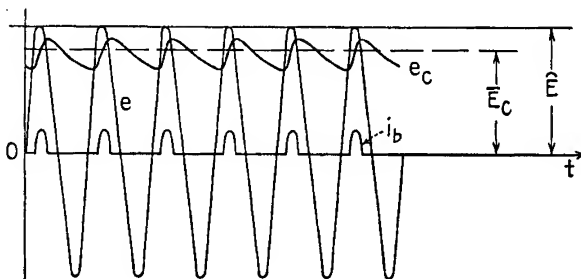


FIG. 4.2.—Showing the voltage across the capacitor and the current in the diode with the RC load of Fig. 4.1.

less than \hat{E} . In Fig. 4.2 the magnitudes of e_c and i_b are sketched for conditions where the d-c voltage developed across the load is approximately 0.8 times the amplitude of the applied r-f voltage. There is current through the diode during only a small fraction of the positive half cycles of the radio-frequency voltage. Each pulse of plate current begins shortly before and ends shortly after the applied high-frequency voltage has its maximum positive value.

As a result of this action there is a direct current in the circuit, which must flow through the resistance R since the average current through C must be zero in the steady state. This average current equals the average plate current \bar{I}_b of the diode. The average voltage across R and across C are identical, so that $\bar{E}_C = \bar{I}_b R$. As before, the polarity of the voltage across the load is such that the plate is biased negatively with respect to the cathode, and

$$\bar{E}_b = -\bar{I}_b R \quad (4.1)$$

The diode load is sometimes arranged as in Fig. 4.3, where the resistor R is in shunt with the diode and the capacitor C is in series

with this combination. With the same R and C and the same sinusoidal applied voltage, this circuit will develop the same average voltage across R provided that the r-f reactance of C is negligibly small. The average voltage across R appears also across C with the polarity shown. With the same r-f applied voltage and the same average voltage across C , the diode has impressed upon it in Fig. 4.3 the same voltages as in Fig. 4.1 and will conduct identically; therefore, the average currents will be the same in both circuits. The average current in the diode and in the resistor R (\bar{I}_b and \bar{I}_R , Fig. 4.3) must be equal in magnitude since the average current through C must be zero. Their directions are shown in the diagram.

The arrangement of Fig. 4.3 is sometimes employed in radio receivers and in vacuum-tube voltmeters. In transmitter and oscillator circuits a similar connection is often employed to obtain grid-bias voltage, the grid serving as the rectifying element and the average voltage across R being used as the grid-bias voltage.

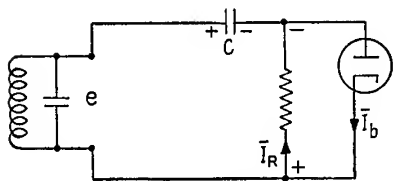


FIG. 4.3.—Alternative arrangement of RC diode load. Signs indicate polarity of steady, or d-c, voltages.

The load on the driving source in Fig. 4.3 is somewhat greater than in Fig. 4.1, because, in addition to the same $\bar{I}^2 R$ due to the rectified d-c, there is in Fig. 4.3 additional power amounting to $\hat{E}^2/2R$ watts dissipated in R due to the r-f voltage applied across R through the capacitor C . The input resistance R_{in} may be evaluated by determining the total power dissipated. In Fig. 4.1, the power dissipated in R due to direct current is $\bar{I}_b^2 R$. Neglecting the power dissipated at the plate of the diode and in R due to the small r-f current caused by the zigzag fluctuations in e_c (the r-f “ripple” voltage), the power input must equal the power dissipated in R , or

$$\frac{\hat{E}^2}{2R_{in}} = \bar{I}_b^2 R = \frac{\bar{E}_c^2}{R}$$

If R is very large, $\bar{E}_c \doteq \hat{E}$, and

$$R_{in} \doteq \frac{R}{2}$$

The smaller the value of R , the smaller the value of \bar{E}_c and the smaller the value of R_{in} .

In the circuit of Fig. 4.3, the power input is equal to the dis-

sipation in R (with the same approximations as above), so that

$$\frac{\hat{E}^2}{2R_{in}} \doteq \bar{I}_b^2 R + \frac{\hat{E}^2}{2R} = \frac{\bar{E}_c^2}{R} + \frac{\hat{E}^2}{2R}$$

If R is very large, $\bar{E}_c \doteq \hat{E}$, and

$$\frac{\hat{E}^2}{2R_{in}} \doteq \frac{\hat{E}^2}{R} + \frac{\hat{E}^2}{2R} = \frac{3\hat{E}^2}{2R}$$

Therefore,

$$R_{in} \doteq \frac{R}{3}$$

If the applied r-f voltage is modulated, the tendency to charge C to the peak value still persists, with the result that when the r-f

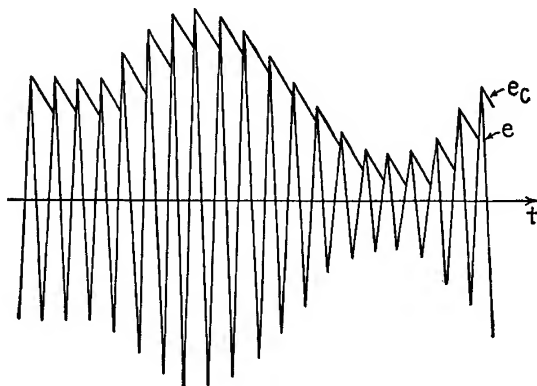


FIG. 4.4.—Fluctuations of e_c in the circuit of Fig. 4.1 when the applied r-f voltage is modulated.

voltage rises during modulation the d-c voltage developed across C also rises. When the r-f voltage falls during modulation, the charge on C must diminish sufficiently rapidly between r-f peaks to enable the voltage across the load to follow the decrease in the crest values of the r-f voltage. Figure 4.4 compares the magnitude of the instantaneous voltage across the diode load of Fig. 4.1 with the value of a sinusoidally modulated applied signal (under correct conditions of operation). The voltage across the load varies in synchronism with the modulation so that the circuit produces the desired audio- or video-frequency output voltage. This voltage is usually applied to following amplifier stages in the receiver.

5. Rectification Diagram.—The quantitative relationships among the amplitude of the applied r-f voltage \hat{E} , the average recti-

fied current \bar{I}_b , and the average anode voltage \bar{E}_b (which equals the average voltage across the load) may be studied by means of a family of curves obtained as follows: In the circuit of Fig. 4.1 or 4.3, C is made large enough to by-pass R at the carrier frequency or to minimize the fluctuations in the instantaneous voltage across C . With the applied voltage held at a constant amplitude \hat{E} , R is varied, causing \bar{I}_b and \bar{E}_b to vary. \bar{I}_b is measured, and \bar{E}_b is calculated from (4.1) or is measured.¹ The variations in \bar{I}_b and \bar{E}_b are plotted, \bar{E}_b horizontally and \bar{I}_b vertically, as in Fig. 6.1. One such curve is obtained for each value of \hat{E} . By suitable choice of \hat{E} a family of curves is obtained. In tube handbooks, \bar{I}_b is described usually as "rectified microamperes" and \bar{E}_b as "d-c volts developed by diode."² With a fixed value of \hat{E} , decreasing the value of the diode-load resistance decreases \bar{E}_b , the limit being zero for zero resistance. As the diode-load resistance is increased, the value of \bar{E}_b increases in the negative direction, and for the diode characteristics assumed in Sec. 3 the upper limit is $\bar{E}_b = \hat{E}$ when the load resistance is infinite.

The family of curves on the rectification diagram is similar geometrically to the family of i_b - e_c curves for a triode. As in the triode amplifier, the audio- or video-frequency voltage across the diode load may be calculated graphically from the characteristic curves or analytically from an equivalent plate circuit. These methods will be explained in turn.

6. Graphical Solution.—In amplifier circuits, distinction has to be made between the d-c and the a-c (audio- or video-frequency) values of the load impedance. The same distinction is equally important in detector circuits. Since an r-f by-pass capacitor C , Fig. 4.1, is an essential part of the circuit, it may also have appreciable shunting effect at audio and video frequencies. However, if the r-f carrier frequency is high, C may be made small enough to

¹ Since $\bar{E}_b = -|\bar{E}_c|$ is really a biasing voltage to the diode, the RC load for purposes of measurement could be replaced by a d-c voltage source of negligible internal impedance and adjustable voltage, with the positive terminal connected to the cathode. Also, for purposes of measurement, a low frequency could be used as the carrier frequency, the difficulties of measurement at radio frequencies being thus avoided.

² RCA Receiving Tube Manual, Technical Series RC-14, pp. 96, 112, RCA Manufacturing Co., Inc. In the tube handbooks the rms r-f voltage is plotted. Since the operation of the circuit is inherently related to the amplitude of the r-f signal, amplitude values are designated in Fig. 6.1 by the rms voltage multiplied by $\sqrt{2}$.

act as an adequate r-f by-pass without appreciably shunting R in the low and middle range of audio frequencies. This is the desirable condition and will be discussed.

In the graphical solution, a line is first drawn from the origin O upward to the left, Fig. 6.1, such that the ratio of \bar{E}_b to \bar{I}_b is equal to R for all points on this line. This line is the d-c load line for the detector circuit and corresponds to the d-c load line employed in the graphical solution of triode problems.¹ This line is the graph

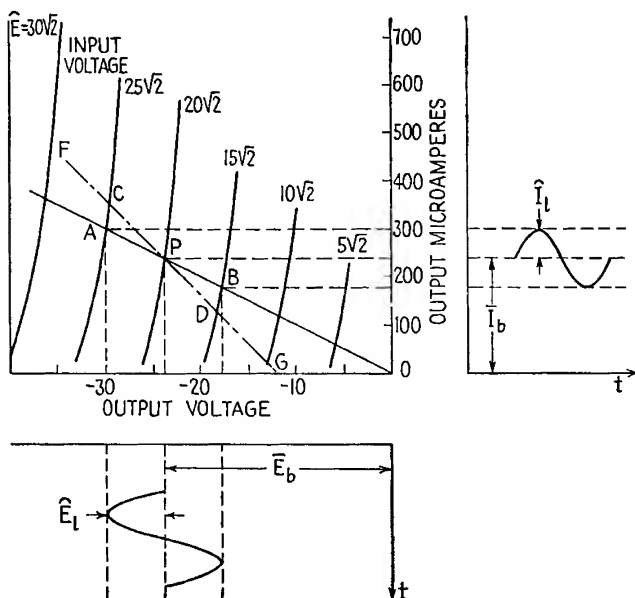


FIG. 6.1.—Diode-rectification diagram showing d-c load line $OBPA$ and a-f resistance load line $FCPDG$. Waveforms of output a-f current I_L and a-f voltage E_L are shown for the condition where the d-c resistance and a-f impedance of the load are equal.

of (4.1), and it runs upward to the left from the origin because in this circuit \bar{E}_{bb} is zero (no polarizing battery used) and the average plate voltage is negative with respect to the cathode. The intersection of this d-c load line with the curve corresponding to the amplitude of the applied r-f voltage determines a point P whose coordinates \bar{I}_b and \bar{E}_b are the direct current through R and the d-c voltage drop across R . This point P might be termed the operating point on the rectification diagram. In Fig. 6.1 the d-c load line is shown for a resistance of 100,000 ohms, and the point P for a carrier

¹ See Chap. XI, Sec. 3. Figs. 3.2 and 3.5.

voltage of $20\sqrt{2}$ volts. The corresponding direct current in the load is 238 microamperes, and the d-c voltage drop across the load is 23.8 volts.

If the voltage applied to the circuit has an amplitude-modulated waveform so that

$$e = \bar{E}(1 + m_a \sin \omega t) \sin \omega_h t \quad (6.1)$$

the point P will move up and down along the d-c load line, provided that at the frequency of modulation the impedance of the load is the same as its resistance to direct current.¹ Then the projection on the horizontal of the point P as it moves up and down the load line traces the variations in E_b , and the vertical projection traces the variations in \bar{I}_b . If the rectification curves are equally spaced,

the variations in \bar{E}_b and \bar{I}_b are proportional to the variations in the amplitude of the applied r-f voltage. An audio- or video-frequency voltage appears across the load. In Fig. 6.1 these variations are sketched for a carrier voltage of $20\sqrt{2}$ volts sinusoidally modulated 25 per cent. At the crest of the modulation cycle, the applied r-f

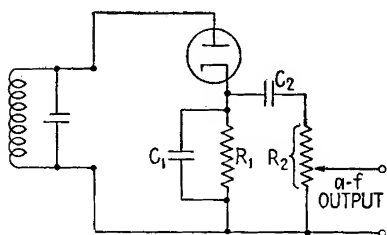


FIG. 6.2.—Showing a diode load whose audio-frequency impedance is smaller than its d-c resistance.

voltage has an amplitude of $25\sqrt{2}$ volts and develops a voltage across the load of -29.8 volts. At the trough of the modulation cycle, the applied r-f voltage is $15\sqrt{2}$ volts (amplitude), and the voltage across the load is -17.8 volts. Under carrier conditions the load voltage is -23.8 volts. Thus the voltage developed across the load (the r-f ripple indicated in Fig. 4.4 being neglected) consists of a d-c component E_b of -23.8 volts and an a-f component \bar{E}_i of 6 volts amplitude. For these conditions the d-c and audio-frequency components of voltage and current in the output are sketched below and to the right of the rectification diagram in Fig. 6.1.

Figure 6.2 shows a circuit in which the diode-load impedance at audio frequencies is less than the d-c resistance. The resistor R_1 and capacitor C_1 may have the same values as in the simple RC load of Fig. 4.1. The resistor R_2 is commonly a potentiometer

¹ The point P in Fig. 6.1 is analogous to the Q point of Fig. 3.2, Chap. XI, and the points A and B in Fig. 6.1 are analogous to the points A and B in Fig. 3.2, Chap. XI.

used for manual volume control in a radio receiver. The capacitor C_2 is a blocking capacitor, blocking the d-c voltage across R_1 so that only audio- or video-frequency voltage appears across R_2 . The variations in this impedance as the audio frequency varies are the same as those of the plate-load impedance in a resistance-capacitance-coupled amplifier. C_1 is made large enough to have a low reactance at the carrier frequency but usually is small enough so that its shunting effect at the lower audio frequencies may be neglected. At these frequencies the coupling capacitor C_2 has a reactance comparable with R_2 . In the middle range of audio frequencies, the audio-frequency reactances of C_1 and C_2 may be neglected, so that the effective audio-frequency impedance is a resistance equal to R_1 and R_2 in parallel. At the higher audio frequencies the shunting effect of C_1 reduces the magnitude of the diode-load impedance and causes it to be capacitive. The same considerations governing the lower and upper half-power frequencies of video- and audio-frequency amplifiers are also involved in correctly designing the load impedance of a detector.

The behavior of the circuit of Fig. 6.2 when the modulation frequency is in the middle range of audio or video frequencies may be shown readily with the aid of the rectification diagram. First, the d-c load line is drawn, Fig. 6.1, whose slope corresponds to the d-c resistance of the diode load, R_1 . The intersection of this load line with the curve corresponding to the r-f carrier voltage establishes the d-c value of the voltage developed across the load in the absence of modulation. When modulation is applied, the ratio of audio-frequency voltage across the load to audio-frequency current in the load is not equal to R_1 but in this case is equal to

$$\frac{R_1 R_2}{(R_1 + R_2)},$$

which is the audio-frequency resistance of R_1 and R_2 in parallel. Hence the operating point will move along a line on the diagram that is the *audio-frequency* load line and not the d-c load line.¹ This line is the line CD in Fig. 6.1, where R_1 has been taken as 100,000 ohms and R_2 has been taken as 100,000 ohms. The a-f resistance

$$\frac{R_1 R_2}{(R_1 + R_2)}$$

is 50,000 ohms. The line CD is located in position by laying off

¹ See Chap. XI, Fig. 3.5, for the analogous situation for triode amplifiers.

horizontally to the left from the point P any convenient value of voltage, say 10 volts. From the left extremity of the line so located, measure upward a distance corresponding to 10 volts divided by the a-f resistance, in this case $(10/50,000) = 200$ microamperes. This locates a point such as F . The a-c load line is determined by drawing a line through F and P .

As the input r-f voltage increases and decreases during modulation, the operating point moves along the a-f load line. The upper limit of the path of operation is determined by the intersection of the a-f load line with the characteristic curve corresponding to the maximum value of the amplitude of r-f voltage. Similarly, the lower limit will be determined by the minimum amplitude of the r-f voltage. In Fig. 6.1 the upper limit C and the lower limit D correspond to the same signal as before, *viz.*, a carrier value of $20\sqrt{2}$ volts amplitude, sinusoidally modulated 25 per cent. Comparison of the line CD with the line AB shows that for the same r-f signal the amplitude of the a-f voltage across the load is less (5.5 volts compared with 6 volts) and the amplitude of the a-f current is greater (110 compared with 60 microamperes).

When the a-f load impedance is reactive, the path of operation is elliptical, similar to that shown in Chap. XI, Fig. 3.6, for the triode amplifier. This case is more readily treated by the analytical method discussed in Sec. 10.

When the characteristic curves of the diode detector are not equally spaced or have appreciable curvature, the intercepts along the a-f load line are not equal, and the average current varies with the percentage of modulation. This changes the position of the line CD , usually shifting it upward, its slope remaining unchanged and the average point P moving outward along the d-c load line. This effect is small when the distortion in the output voltage is small and has been neglected for the small percentage of modulation indicated in Fig. 6.1.

7. Distortion and Maximum Allowable Modulation.—With the sinusoidally modulated voltage given by (6.1) maintained across the input terminals, the audio- or video-frequency output voltage may vary from a sinusoidal waveform due to the nonuniform spacing of the characteristics on the rectification diagram. The characteristics are usually nearer together for small than for large r-f voltages. Also, the characteristics have a steeper slope as the rectified or average current increases. If the a-f impedance of the diode load is equal to the d-c resistance, the distortion due to these effects is

small, since the rectification characteristics, as in Fig. 6.1, usually intersect any given d-c load line at nearly equal intervals. However, if the path of operation is along an a-f resistance line, the distortion caused by the curvature increases as the a-f resistance line approaches the vertical, *i.e.*, as the a-f resistance decreases. The distortion due to curvature of the characteristics is also greater for a reactive load having an elliptical path of operation.

When the a-f impedance of the detector load is less than its d-c resistance, there is distortion at high percentages of modulation due to what is called "clipping." The current through the diode cannot reverse; and as the percentage of modulation increases, a limiting condition is reached when the diode current is zero during one or more complete r-f cycles. If the percentage of modulation were increased above that indicated by the line *CD*, Fig. 6.1, the path of operation would extend up to the left of *C* and down to the right of *D*. The limit is reached when the path of operation extends to *G* at the base line of the diagram.¹ For higher percentages of modulation the a-c load line cannot meet the rectification characteristic corresponding to the minimum amplitude of the impressed r-f voltage, and hence the rectified current cannot vary sinusoidally.

Similarly, if the diode load is reactive, an elliptical path of operation results, and the limiting condition is reached when the path of operation becomes tangent to the base line on the diagram. This condition is encountered commonly at high frequencies of modulation, where the capacitor *C* of the diode load of Fig. 4.1 has appreciable shunting effect at the higher audio or video frequencies. The time constant *RC* of the diode load then is appreciable compared with the time of the modulation cycle, and the discharging action pictured in Fig. 4.4 cannot take place rapidly enough to allow the average voltage (over each succeeding r-f cycle) to decrease fast enough to follow the decline in the envelope of the modulated wave.

Figure 7.1 is a cathode-ray oscillogram of the voltage across an *RC* detector load as connected in Fig. 4.1 and shows the desired conditions of operation. Proper charging and discharging action is shown as sketched in Fig. 4.4. The upward trace corresponding to

¹ This limit can be extended by applying to the plate a positive bias that varies with the carrier voltage. See *Radionics*, 74, 21, March, 1937, or "Radiotron Designers Handbook," 3d ed., p. 163, Wireless Press for Amalgamated Wireless Valve Co. Pty. Ltd., distributed in U.S.A. by RCA Manufacturing Co., Inc., 1941, edited by F. Langford Smith.

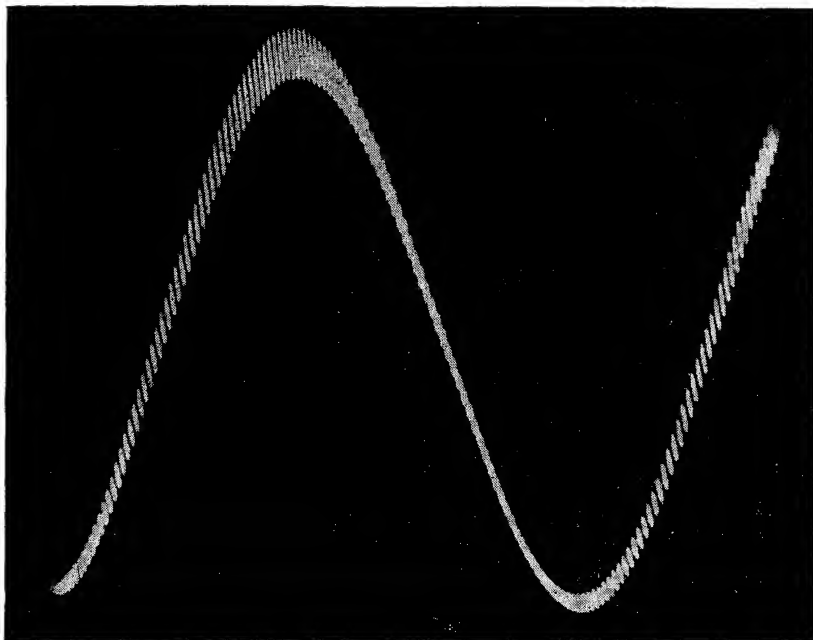


FIG. 7.1.—Cathode-ray oscillogram of the voltage across an *RC* load showing normal operation.

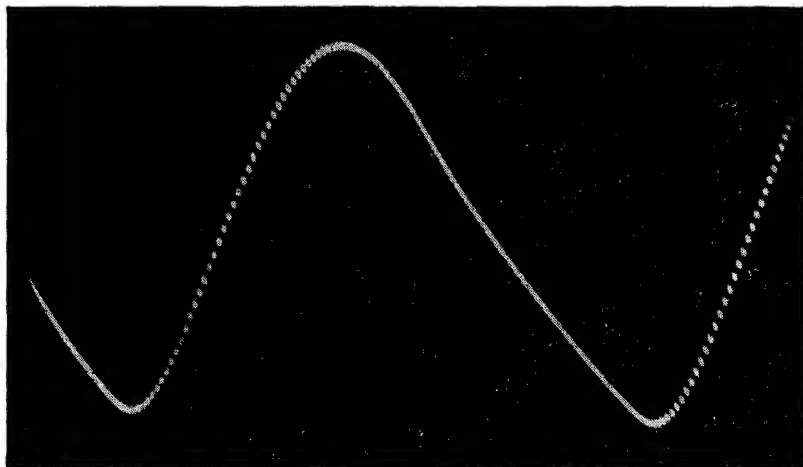


FIG. 7.2.—Oscillogram of the voltage across an *RC* load having too large a time constant, causing diagonal clipping.

the charging part of the r-f cycle is absent in the photograph; because the high speed of movement of the spot on the cathode-ray screen caused the trace to be so dim that it was hardly visible. The slower and brighter discharge traces are clearly shown. The low-frequency component of the voltage across the load is sinusoidal.

Figure 7.2 shows an oscillogram of the type of distortion encountered when the capacitor C of the diode load of Fig. 4.1 is too large. The discharge of C through R is so slow that the sides of the downward trace of Fig. 7.1 are clipped. The path of operation on the rectification diagram is elliptical, with a section cut off by the base of the diagram. Figure 7.3 shows an oscillogram of the type of

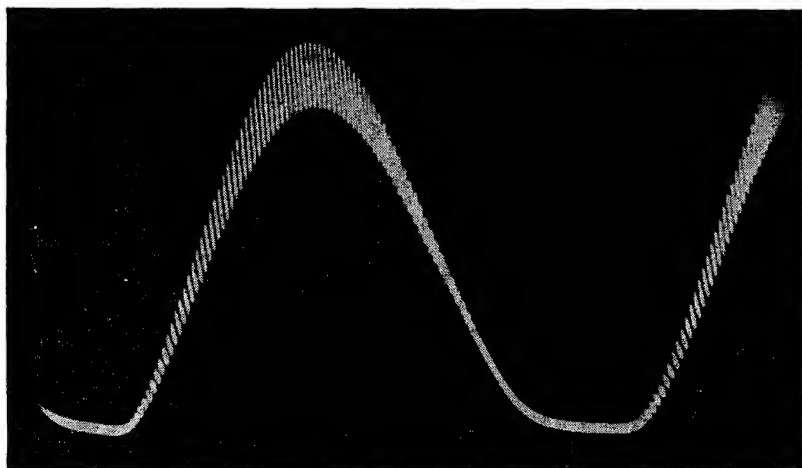


FIG. 7.3.—Oscillogram of the voltage across a load whose a-f resistance is considerably less than the d-c resistance, showing bottom clipping.

distortion encountered in a circuit such as that of Fig. 6.2, in the middle range of audio frequencies at which the a-f resistance is considerably less than the d-c resistance. The bottom of the desired sinusoid is clipped. The path of operation on the rectification diagram is a line such as $FCDG$ in Fig. 6.1, with the modulation greater than the allowable limit.

The diagonal and bottom clipping shown in Figs. 7.2 and 7.3 results when the percentage of modulation of the applied voltage exceeds that which the circuit is capable of detecting linearly. The limit is reached when the amplitude of the low-frequency (audio- or video-frequency) component of the output current, shown as \hat{I}_l on the right of Fig. 6.1, equals the d-c component \bar{I}_b . The magnitude of \hat{I}_l cannot be greater than this without clipping because the

current through the diode cannot reverse. In the limit, $|\hat{I}_l| = \bar{I}_b$. The corresponding a-f and d-c voltages across the diode load (across R_l in Fig. 6.1) are

$$|\hat{E}_l| = |\hat{I}_l Z_l| \quad (7.1)$$

and

$$|\bar{E}_{b1}| = |\bar{I}_b R_l| \quad (7.2)$$

The limiting ratio of a-f to d-c load voltage is

$$\frac{|\hat{E}_l|}{|\bar{E}_{b1}|} = \frac{|\hat{I}_l Z_l|}{|\bar{I}_b R_l|} = \left[\frac{|Z_l|}{|R_l|} \right]_{\hat{I}_l = \bar{I}_b}^{\text{with}} \quad (7.3)$$

The d-c output voltage $|\bar{E}_b|$ cannot be varied by a greater percentage than that indicated by (7.3) without clipping; and when R_l and Z_l are large,¹ the input carrier cannot be modulated to a greater percentage than this because the envelope will not be reproduced in the output.

Hence, even with straight and uniformly spaced rectification characteristics, the diode-detector circuit cannot detect without clipping when the modulation exceeds the approximate value

$$m_a \Big|_{\max} = \frac{|Z_l|}{R_{dc}}$$

where $|Z_l|$ is the magnitude of the low-frequency impedance of the diode load and R_{dc} is its resistance to the rectified direct current. To avoid clipping, the diode load should be designed so that its impedance at the modulation frequency is as nearly equal to the d-c load resistance as the required circuit performance will permit.

8. Automatic Volume Control.—The d-c voltage developed across the diode load may be used to control the gain of amplifier or converter stages in a radio receiver, by varying the bias voltage on the grids of the tubes. Figure 8.1 is, with slight changes, part of the wiring diagram of a broadcast receiver² showing a diode detector, with automatic volume control (AVC) applied to the intermediate-frequency amplifier tube preceding the detector. (In the receiver from which this diagram was taken and in other receivers, AVC voltage is applied to more than one stage.) The time constant $R_{12}C_{15}$ is of the order of 0.1 sec, with R_{12} usually about 1 megohm in order to keep the a-f impedance of the diode load high.

¹ Compared with r_d as defined in Sec. 10.

² See RCA Receiving Tube Manual, Technical Series RC-14, p. 204, RCA Manufacturing Co., Inc., 1940; also Fig. 5.2, Chap. XXIII.

C_{15} is of the order of $0.1 \mu\text{f}$, so that its reactance at audio and radio frequencies is small compared with R_{12} . Very little a-f or r-f voltage appears across C_{15} , the combination of R_{12} and C_{15} acting as a "filter," so that the AVC voltage is practically pure d-c.

As the voltage derived from the receiving antenna increases, the voltage applied to the diode tends to increase, thus increasing the d-c voltage across the diode load and across C_{15} . This d-c voltage is fed to the amplifier grids, through decoupling filters such as R_8C_{11} , decreasing the g_m of the tubes and reducing the gain. The result is that the voltage applied to the diode does not increase as much as the input voltage to the receiver and can be maintained approximately constant except when the input voltage is very small.

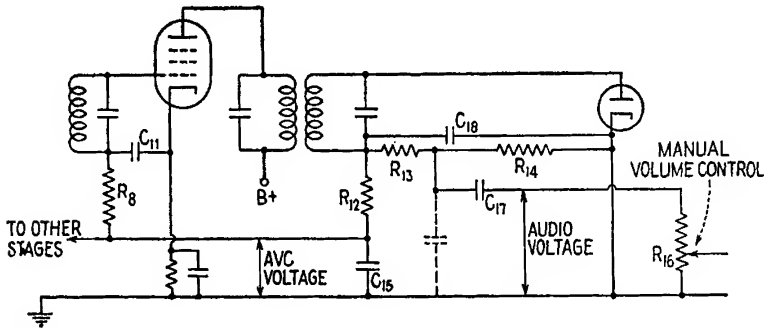


FIG. 8.1.—Detector circuit with automatic volume control.

Sometimes a separate diode is used for AVC to avoid loading the diode used for detection, or to modify the AVC action. If AVC is applied to amplifier stages following that to which the AVC diode is connected, it is possible to hold the receiver output voltage constant or even to make it decrease when the input signal voltage exceeds a certain value. AVC is used principally to minimize fluctuations in the output volume of a receiver when the input voltage fluctuates or fades or when the receiver is tuned from a weak station to a stronger one without changing the manual volume control.

9. R-f Filtering.—An r-f choke coil or a low-pass a-c filter circuit may be inserted between the detector and the first low-frequency-amplifier tube to minimize the r-f ripple voltage applied to the grid. Filtering action may also be secured by using RC filter circuits of the decoupling type. In Fig. 8.1 the audio-frequency voltage is obtained from a tap on the diode-load resistance,

i.e., from the connection between R_{13} and R_{14} . R_{13} together with the distributed capacitance of the audio-frequency circuit acts as a filter, reducing the amount of r-f voltage across R_{16} . Sometimes this action is augmented by a capacitor added to the circuit in the position shown dotted in Fig. 8.1. Such additions decrease the low-frequency impedance of the diode load and cause clipping to occur at smaller percentages of modulation.

10. Analytical Solution, Equivalent Circuit.—Where the diode load is appreciably reactive at the modulation frequency, the path of operation on the rectification diagram becomes elliptical. Constructing this elliptical path on the rectification diagram is not easy, while the analytical method is relatively simple. This analytical method is practically the same as the equivalent-plate-circuit method for amplifiers. Comparison of the rectification diagram in Fig. 6.1 with the family of plate-current-grid-voltage curves of a triode shows that they are geometrically similar. For triodes, the plate current can be expressed as

$$i_b = f(e_b, e_c)$$

Here, \bar{I}_b can be expressed as

$$\bar{I}_b = f(\bar{E}_b, \hat{E}) \quad (10.1)$$

where \hat{E} is the amplitude of the applied sinusoidal r-f voltage. Hence,

$$d\bar{I}_b = \frac{\partial \bar{I}_b}{\partial \bar{E}_b} d\bar{E}_b + \frac{\partial \bar{I}_b}{\partial \hat{E}} d\hat{E} \quad (10.2)$$

Just as $\partial e_b / \partial i_b$ is defined as the plate resistance r_p of a triode, so $\partial \bar{E}_b / \partial \bar{I}_b$ can be defined as the plate resistance r_d for detection or, simply, the detection plate resistance. Then

$$r_d = \left. \frac{d\bar{E}_b}{d\bar{I}_b} \right]_{\hat{E} \text{ const.}} \quad (10.3)$$

If \bar{I}_b is kept constant, (10.2) becomes

$$0 = \frac{\partial \bar{I}_b}{\partial \bar{E}_b} d\bar{E}_b + \frac{\partial \bar{I}_b}{\partial \hat{E}} d\hat{E}$$

so that

$$-\left. \frac{d\bar{E}_b}{d\hat{E}} \right]_{\bar{I}_b \text{ const.}} = + \left(\frac{\partial \bar{I}_b}{\partial \hat{E}} \right) \left(\frac{\partial \bar{E}_b}{\partial \bar{I}_b} \right)$$

This ratio is analogous to the μ of a triode, so that the “detection

amplification factor μ_d ," or "detection μ ," may be defined as

$$\mu_d = - \left. \frac{d\bar{E}_b}{d\bar{E}} \right]_{I_b \text{ const.}} = - \frac{\partial \bar{E}_b}{\partial \bar{E}} \quad (10.4)$$

Referring to Fig. 6.1, and considering amplitude values of the incremental quantities indicated in (10.2) when the applied voltage is sinusoidally modulated,

$$d\bar{I}_b = \hat{I}_i \quad \text{and} \quad d\bar{E} = m_a \hat{E}$$

and

$$d\bar{E}_b = (-d\bar{I}_b)Z_L = -\hat{I}_i Z_L$$

where Z_L is the low-frequency impedance of the diode load. Equation (10.2) becomes

$$\hat{I}_i = \frac{1}{r_d} (-\hat{I}_i Z_L) + \frac{\mu_d}{r_d} m_a \hat{E} \quad (10.5)$$

whence

$$\hat{I}_i = \frac{\mu_d m_a \hat{E}}{r_d + Z_L} = \frac{\hat{E}_i}{r_d + Z_L} \quad (10.6)$$

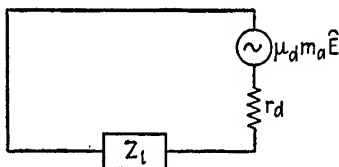


FIG. 10.1.—Equivalent low-frequency circuit for linear detection.

where \hat{E}_i is the generator voltage in the equivalent plate circuit of Fig. 10.1. The audio- or video-frequency voltage across the load (across C_1 , Fig. 6.2, or across C_{18} , Fig. 8.1) is

$$\hat{V}_i = \mu_d m_a \hat{E} \frac{Z_L}{r_d + Z_L} \quad (10.7)$$

Similar equations hold for rms values. Equation (10.6) is analogous to (9.7), of Chap. XI, and is the equation for the low-frequency current in the equivalent low-frequency circuit of Fig. 10.1. By means of this circuit the audio- or video-frequency voltage across the detector load can be calculated for any impedance when the detection amplification factor and detection plate resistance have been evaluated. For the 6H6 diode, μ_d is 0.93 and r_d is 14,000 ohms, approximately. As with triodes these values are not constant because the characteristic curves are not perfectly straight or perfectly spaced.

SMALL-SIGNAL (SQUARE-LAW) DIODE DETECTION

11. Difference between Linear and Square-law Diode Detection.—Linear detection by a diode-detector circuit requires a large

applied voltage; the diode plate is driven negative with respect to the cathode, and the diode does not conduct during most of the r-f cycle. On the other hand, square-law operation takes place when the driving voltage is so small that conduction never ceases during the r-f cycle. The variations in the plate current of the diode then resemble the variations in the plate current of a Class A triode amplifier operated in a region where the tube characteristics

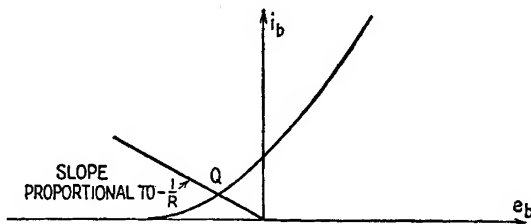


FIG. 11.1.—Current-voltage characteristic of a diode in the vicinity of zero plate voltage, showing the Q point of the diode in the circuit of Fig. 4.1.

are very curved. In Fig. 11.1 a typical diode i_b - e_b characteristic in the vicinity of zero plate voltage is sketched to an expanded vertical scale. There is some current even when the plate voltage is zero, as described in Sec. 3. For small voltages the diode acts, not as a one-way rectifier, but as a finite, nonlinear series resistance. The Q point on the i_b - e_b characteristic of the diode can be determined by the method outlined in connection with Fig. 3.5, Chap. X. If a circuit such as that of Fig. 4.1 is used, there is no external source of bias voltage and the d-c load line is drawn from the origin upward to the left as in Fig. 11.1. In the following paragraphs, an external polarizing source will be assumed, for generality.

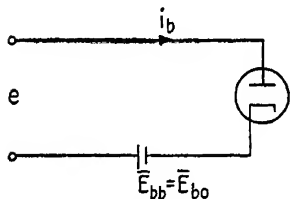


FIG. 12.1.—Diode detection with no load.

12. Square-law Rectification.—Figure 12.1 shows a circuit in which a diode is connected in series with a polarizing battery assumed to have zero internal impedance. The driving source is assumed to maintain a sinusoidal voltage e across the input terminals. The diode is assumed to have the nonlinear characteristic shown in Fig. 12.2. When the r-f voltage is applied, the current increase during the positive half cycle is greater than the decrease during the negative half cycle, Fig. 12.2, with the result that the average current increases. This increase is the useful indication of the presence of an r-f voltage.

The diode is assumed to have the nonlinear characteristic shown in Fig. 12.2. When the r-f voltage is applied, the current increase during the positive half cycle is greater than the decrease during the negative half cycle, Fig. 12.2, with the result that the average current increases. This increase is the useful indication of the presence of an r-f voltage.

If the high-frequency (r-f) applied voltage is

$$e = \bar{E} \sin \omega t \quad (12.1)$$

the current variations about the Q point may be represented by a Taylor's series,

$$i_{r0} = \frac{\partial i_b}{\partial e_b} e + \frac{1}{2!} \frac{\partial^2 i_b}{\partial e_b^2} e^2 + \frac{1}{3!} \frac{\partial^3 i_b}{\partial e_b^3} e^3 + \cdots \quad (12.2)$$

the derivatives being those at the Q point. Substitution of (12.1) in (12.2) and expansion yield terms that are d-c, or constant, terms

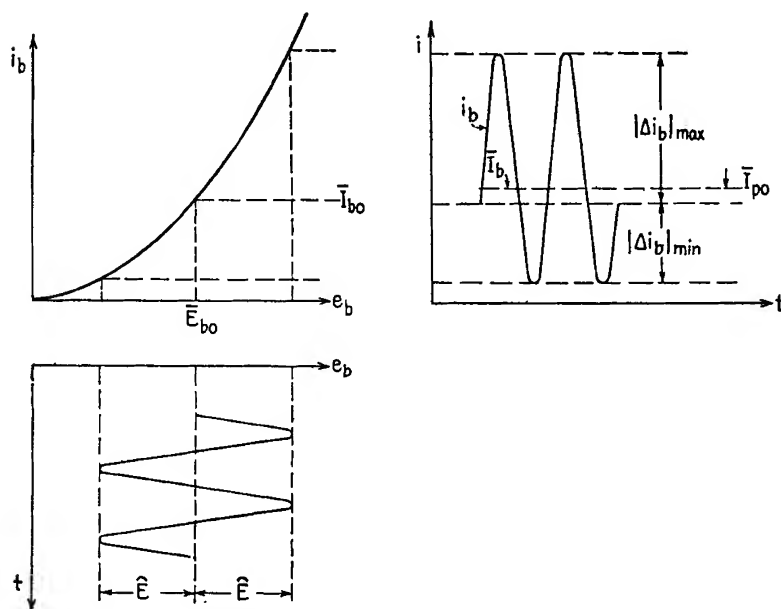


FIG. 12.2.—Input voltage and output current of unloaded square-law detector.

and other terms involving the fundamental radio frequency and its harmonics. For purposes of detection, the d-c or constant terms are the useful terms, accounting for the increase of \bar{I}_b over \bar{I}_{b0} in Fig. 12.2. Of all the terms on the right-hand side of (12.2), the second derivative term contributes the greatest amount to the increase in average current. The first and third derivative terms contribute nothing. The fourth and higher derivative terms contribute very little. Therefore, derivatives higher than the second need not be considered, and an ideal case will be assumed in which the higher order terms are zero. Then (12.2) becomes

$$i_{p0} = \frac{\partial i_b}{\partial e_b} \hat{E} \sin \omega_n t + \frac{1}{4} \frac{\partial^2 i_b}{\partial e_b^2} \hat{E}^2 - \frac{1}{4} \frac{\partial^2 i_b}{\partial e_b^2} \hat{E}^2 \cos 2\omega_n t \quad (12.3)$$

The average increase in current is given by the second term on the right of (12.3). When denoting this by \bar{I}_{p0} , the difference between \bar{I}_b and \bar{I}_{b0} in Fig. 12.2 is

$$\bar{I}_{p0} = \frac{1}{4} \frac{d^2 i_b}{de_b^2} \hat{E}^2 \quad (12.4)$$

the derivative to be evaluated at the Q point. This equation shows that the increase in average current is proportional to the square of the amplitude of the r-f voltage. Further, the increase depends upon the second derivative of the current, or upon the curvature or rate of change of the slope of the diode characteristic. The greater the curvature, the more sensitive the detection.

For a parabolic characteristic, for example $i_b = Ke_b^2$, where K is a constant, the second derivative $d^2 i_b / de_b^2 = 2K$, and \bar{I}_{p0} is independent of the Q point. Also, if $|\Delta i_b|_{\max}$ and $|\Delta i_b|_{\min}$ are denoted as in Fig. 12.2,

$$\bar{I}_{p0} = \frac{|\Delta i_b|_{\max} - |\Delta i_b|_{\min}}{4}$$

If in practice a parabolic characteristic is not obtained, the rectifying device may be polarized to place the Q point where the curvature is greatest.

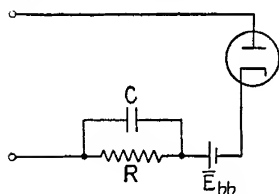


FIG. 13.1.—Diode detector with by-passed resistive load.

13. Square-law Detection, By-passed Resistive Load.—Some form of load must be connected in series with the diode in order to use the rectified current. One of the simplest and most useful loads is a resistor by-passed for the radio frequencies by a capacitor, Fig. 13.1. With the capacitance sufficiently large, practically all

the r-f voltage applied to the circuit is impressed on the diode. For the same Q point as in Sec. 12, \bar{E}_{bb} must be increased. As here indicated in Fig. 13.2 the Q point and the value of \bar{E}_{bb} are connected on the i_b - e_b diagram by the d-c load line as explained in Chap. X. When the radio-frequency voltage is applied to the circuit, the diode will rectify in a manner similar to that shown in Fig. 12.2, but owing to the presence of R the action of the diode tending to increase the d-c, or average, current will be opposed by the resistance R . For this reason, the increase in average current for the same impressed

r-f voltage will be smaller than in the unloaded detector. The d-c load line and the line representing the average current \bar{I}_b intersect at the A point, Fig. 13.2. The voltage \bar{E}_b at A is the average plate potential when the r-f signal is applied, the rectification of the signal causing the average plate current to increase and the average plate voltage to decrease from the values at the Q point. The a-c

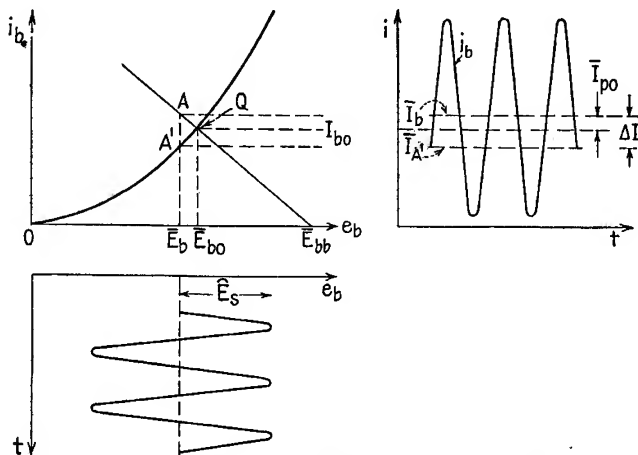


FIG. 13.2.—Current and voltage in the circuit of Fig. 13.1.

components of plate voltage fluctuate about the point A' , Fig. 13.2, and the total voltage impressed on the diode is

$$\begin{aligned} e_b &= \bar{E}_{bb} - \bar{I}_b R + \hat{E} \sin \omega_h t \\ &= \bar{E}_b + \hat{E} \sin \omega_h t \end{aligned} \quad (13.1)$$

So far as the diode is concerned, the voltage conditions with respect to the point A' are the same as in Sec. 12 with respect to the Q point. Therefore, the increase in average current measured from A' must be the same as in Sec. 12. This increase is given by (12.4) and for a parabolic characteristic is independent of the operating point. Therefore,

$$\bar{I}_b - \bar{I}_{A'} = \frac{1}{4} \frac{\partial^2 i_b}{\partial e_b^2} \hat{E}^2 = \Delta I \quad (13.2)$$

as in Fig. 13.2. The increase in average current resulting from the applied r-f voltage is

$$\bar{I}_{p0} = \bar{I}_b - \bar{I}_{b0} \quad (13.3)$$

which, after substituting the value of \bar{I}_b from (13.2), becomes

$$\bar{I}_{p0} = \Delta I - (\bar{I}_{b0} - \bar{I}_{A'}) \quad (13.4)$$

Assuming the points A' and Q to be so close together that the slope of the i_b - e_b characteristic may be considered constant in that interval,

$$\begin{aligned} \bar{I}_{b0} - \bar{I}_{A'} &= \frac{\partial i_b}{\partial e_b} (\bar{E}_{b0} - \bar{E}_b) \\ &= \frac{1}{r_p} \bar{I}_{p0} R \end{aligned} \quad (13.5)$$

where r_p is the variational, or a-c, plate resistance of the diode at the Q point. Substituting (13.5) and (13.2) in (13.4),

$$\bar{I}_{p0} = \frac{\frac{r_p}{4} \frac{\partial^2 i_b}{\partial e_b^2} \bar{E}^2}{r_p + R} \quad (13.6)$$

This is the increase in current above the value at the Q point caused by the r-f signal of amplitude \bar{E} .

14. Equivalent Circuit.—The numerator of (13.6) has the dimensions of voltage and the denominator the dimensions of resistance;

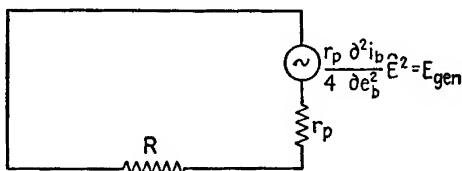


FIG. 14.1.—Equivalent circuit for calculating results of square-law detection.

hence (13.6) is the equivalent of Ohm's law for the results of square-law detection. For calculating the increase in current, the circuit of Fig. 13.1 may be replaced by the equivalent circuit of Fig. 14.1. The diode is replaced by an equivalent generator, or source, of incremental direct current, whose generated voltage E_{gen} is

$$\left(\frac{r_p}{4} \right) \left(\frac{d^2 i_b}{d e_b^2} \right) \bar{E}^2$$

and whose internal impedance is r_p . The load resistance is R . Basically this circuit is the same as the equivalent circuits for amplifiers and large-signal detectors. The parameters differ, but the method of calculation is the same. The equivalent voltage is a property of the tube and is independent of the load. Note that for

square-law detection the variational current resulting from rectification varies as the square of the amplitude of the r-f voltage across the rectifier.

15. Square-law Detection, Amplitude-modulated Signal.—It is shown in Secs. 12 and 13 that, when an unmodulated wave is impressed upon a square-law detector, there is an increase in average current proportional to the square of the amplitude of the r-f voltage. If the amplitude of the r-f voltage is modulated, the increase in average current is also modulated, or changed, in proportion to the square of the amplitude of the r-f voltage. This modulation, or change, in the average current causes audio- or video-frequency currents in the detector output.

Before calculating the low-frequency currents in the detector output, it may be well to investigate the question of what frequencies exist in the current through a square-law impedance when an amplitude-modulated wave is impressed. If the impressed voltage is sinusoidally amplitude-modulated, e is given by

$$e = [\hat{E}(1 + m_a \sin \omega t)] \sin \omega_h t \quad (15.1)$$

where the modulated amplitude of the wave is the expression within the brackets. This wave consists of three frequencies, as is shown by expanding (15.1) to give

$$e = \hat{E} \sin \omega_h t - \frac{m_a}{2} \hat{E} \cos (\omega_h + \omega_i)t + \frac{m_a}{2} \hat{E} \cos (\omega_h - \omega_i)t \quad (15.2)$$

When the square-law modulator was discussed, Chap. XX, Sec. 6, the basic equation was the same as (12.2). It was shown that, when voltages of more than one frequency are impressed on a square-law impedance, the first-power term yields currents having the same frequencies as the input voltages; the second-power term yields the second harmonics of the input frequencies together with frequencies that are the sum and difference of the impressed frequencies. With this general principle, it may be stated that, in response to the three frequencies of (15.2), the current through the unloaded square-law detector, will have the input frequencies, the second harmonic frequencies, and the sum and difference frequencies taken two at a time, as tabulated in Table 15.1.

In audio-frequency transmission, the modulation is not sinusoidal or pure-tone modulation; a more or less broad band of frequencies is involved. The upper and lower side bands each contain the same number of frequencies and the process of detection

produces their second harmonics and sum and difference frequencies. The presence of more than one frequency in each side band causes the square-law-detector output to have frequencies equal to the difference between the side-band frequencies within each of the upper and lower side bands, so that new audio frequencies appear which are the sum and difference of the original audio tones in the audio signal at the transmitter. These terms are not accounted for in Table 15.1, although the three main categories of the output frequencies are clearly indicated.

TABLE 15.1.—FREQUENCIES INVOLVED IN SQUARE-LAW DETECTION

Frequencies of a-c output		Source
Input frequencies	ω_h	Carrier
	$\omega_h + \omega_l$	Upper side band (usb)
	$\omega_h - \omega_l$	Lower side band (lsb)
Second harmonics	$2\omega_h$	Square of carrier
	$2\omega_h + 2\omega_l$	Square of usb
	$2\omega_h - 2\omega_l$	Square of lsb
Sum and difference frequencies	$2\omega_h + \omega_l$	Carrier + usb
	$2\omega_h - \omega_l$	Carrier + lsb
	$2\omega_h$	Usb + lsb
	ω_l	Usb - carrier
	ω_l	Carrier - lsb
	$2\omega_l$	Usb - lsb

The operation of the diode detector without load may be expressed graphically as in Fig. 15.1. The output-current wave is distorted considerably when compared with the input-voltage wave and contains all the frequencies of Table 15.1. The low-frequency variation of the current is shown by the dashed line, and this fluctuation is a sort of varying average over each successive r-f cycle. The low-frequency variation, as measured from the quiescent value, involves a steady component, similar to (and slightly greater than) \bar{I}_{p0} of Sec. 13, and the low-frequency components of angular frequencies ω_l and $2\omega_l$ of Table 15.1. The audio reproducing equipment responds to the low-frequency components.

The desired frequency in the detector output is the frequency with which the r-f wave is modulated; the desired terms in Table 15.1 are the ω_l terms. The $2\omega_l$ term cannot be neglected nor can it always be filtered out, because the band of frequencies required for audio reproduction is many octaves wide, and the second har-

monics of all but the highest tones, or frequencies, fall within the band. In the unloaded detector the strength of the ω_l and $2\omega_l$ terms in the detector output can be obtained by substitution of (15.2) in (12.2). This procedure does not give the answer for the detector with load. The problems of the loaded and unloaded detector can be solved by using the principle developed in Secs. 13 and 14. It is shown in (13.6) that the incremental current resulting from

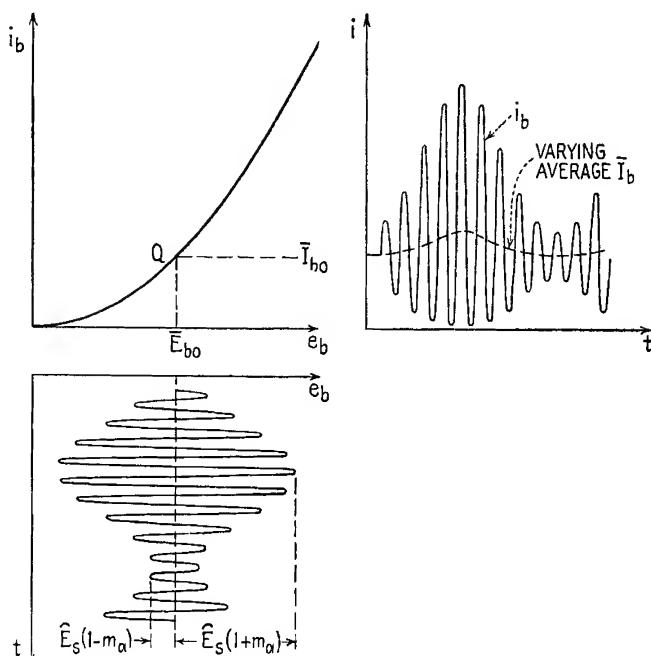


FIG. 15.1.—Operation of diode square-law detector with no load, modulated voltage impressed.

detection is obtained by dividing the equivalent voltage of detection (proportional to the square of the amplitude of the r-f voltage) by the resistance of the circuit to the incremental current. This principle may be extended to cover the case where the amplitude of the r-f voltage is modulated. Since the amplitude of the r-f voltage is now the expression within the brackets in (15.1), the equivalent emf in the equivalent circuit of Fig. 14.1 becomes

$$\frac{r_p}{4} \frac{d^2 i_b}{d e_b^2} [\hat{E}(1 + m_a \sin \omega_l t)]^2$$

or,

$$\begin{aligned} \text{Equivalent emf} &= \frac{r_p}{4} \frac{d^2 i_b}{de_b^2} \hat{E}^2 \left(1 + \frac{m_a^2}{2} \right) \\ &+ \frac{r_p}{4} \frac{d^2 i_b}{de_b^2} \hat{E}^2 (2m_a \sin \omega_l t) \\ &- \frac{r_p}{4} \frac{d^2 i_b}{de_b^2} \hat{E}^2 \left(\frac{m_a^2}{2} \cos 2\omega_l t \right) \end{aligned} \quad (15.3)$$

The $\sin \omega_l t$ term represents the combined effects of the two ω_l terms in Table 15.1, and the $\cos 2\omega_l t$ term represents the $2\omega_l$ term of Table 15.1. If the steady component of the equivalent emf is denoted by \bar{E} and the amplitude of the ω_l component of the equivalent emf is denoted by \hat{E}_l then, from (15.3),

$$\bar{E} = \frac{r_p}{4} \frac{d^2 i_b}{de_b^2} \hat{E}^2 \left(1 + \frac{m_a^2}{2} \right) \quad (15.4)$$

$$\hat{E}_l = \frac{r_p}{4} \frac{d^2 i_b}{de_b^2} \hat{E}^2 (2m_a) \quad (15.5)$$

Similarly, the amplitude of the second harmonic component of the equivalent emf is

$$\hat{E}_{2l} = \frac{r_p}{4} \frac{d^2 i_b}{de_b^2} \hat{E}^2 \left(\frac{m_a^2}{2} \right) \quad (15.6)$$

Note that the ratio of (15.5) to (15.6) is $m_a/4$, so that with pure-resistance load the percentage of second harmonic distortion in a square-law detector is one-fourth the percentage of modulation.

To calculate the desired output current of angular frequency ω_l , the impedances of the circuit at that frequency must be known. Since the output current is an incremental current, the internal impedance of the diode to that current will be r_p , evaluated at the Q point. The impedance of the detector load will depend upon the nature and connection of its elements. If the detector-load impedance at the angular frequency ω_l is denoted by Z_l , then for the circuit of Fig. 13.1, where R and C form the load impedance,

$$Z_l = \frac{R \left(-j \frac{1}{\omega_l C} \right)}{R - j \frac{1}{\omega_l C}}$$

If the amplitude of the desired low-frequency output current is denoted by \hat{I}_l ,

$$\hat{I}_i = \frac{\frac{r_p}{4} \frac{d^2 i_b}{d e_b^2} \hat{E}^2(2m_a)}{r_p + Z_i} = \frac{\hat{E}_i}{r_p + Z_i} \quad (15.7)$$

Equation (15.7) is merely an extension of the equivalent circuit of Fig. 14.1 to the case where the amplitude of the r-f voltage is sinusoidally modulated at the frequency $f_i = \omega_i/2\pi$ cps and is the equation for the equivalent circuit of Fig. 15.2.

16. Detector-circuit Impedances.—

At radio frequencies, the detector-load impedance should be negligible in comparison with the plate resistance of the diode. Then all the r-f voltage impressed on the circuit appears across the diode, and the detector develops the greatest possible low-frequency emf. Also, unwanted r-f voltages across the load, due to any of the r-f currents flowing in the circuit, will be kept to a minimum. This condition is usually secured by using a by-pass capacitor whose r-f reactance is small. The input impedance to the square-law detector circuit is then approximately equal to the r_p of the diode at the Q point.

At audio frequency, *i.e.*, at the frequency of modulation, the impedance of the detector load should be large compared with the internal resistance r_p of the diode in order to have the low-frequency voltage across the load as large as possible. This is shown by the following equation, in which \hat{V}_i is the amplitude of the low-frequency voltage across the load. From Fig. 15.2, $\hat{V}_i = \hat{I}_i Z_i$, and, substituting the value of \hat{I}_i from (15.7),

$$\hat{V}_i = \frac{r_p}{4} \frac{d^2 i_b}{d e_b^2} \hat{E}^2(2m_a) \frac{Z_i}{r_p + Z_i} \quad (16.1)$$

To make Z_i large compared with r_p , the resistance R and the low-frequency reactance of C in Fig. 13.1 should both be large. The latter condition requires a small capacitor. However, this capacitor should not be so small that it cannot act as an r-f by-pass around R .

The source supplying the r-f voltage should have negligible impedance to audio-frequency currents. This condition usually is attained in practice, since the diode with its series load is connected commonly across a parallel tuned circuit whose coil has negligible impedance at low frequencies.

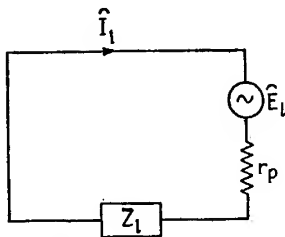


FIG. 15.2.—Equivalent plate circuit for fundamental audio-frequency current resulting from detection of a modulated wave.

17. Selective Fading.—In voice transmission, the radio wave contains many side-band frequencies, of which the single pair considered in Table 15.1 is an example. If the wave at the terminals of the receiving antenna is an undistorted copy of that at the terminals of the transmitting antenna, the two ω_i terms of Table 15.1 are in phase and add to make the desired audio-frequency signal in the detector output.¹ If ionospheric or atmospheric conditions cause the transmission of the radio wave to have phase distortion, the alteration in the phases of the side-band frequencies will cause the two ω_i terms to be more or less out of phase and their sum will vary, causing the tone corresponding to ω_i to be more or less weakened. These phase shifts occur sporadically and may cause individual tones in the detector output to fade out and in, even though the received wave as a whole remains strong. This phenomenon is known as selective fading and is responsible for the peculiar quality frequently observed in voice transmission over long distances. The same effect occurs with linear detection, since phase distortion in transmission causes the received wave to have an envelope different from the wave at the transmitter.

18. Carrier and Side-band Suppression.—The presence of two ω_i terms in Table 15.1, one of them derived from the upper side band and carrier and the other from the lower side band and carrier, indicates that both side bands are not essential to the generation of the low frequency in the detector circuit. One side band and the carrier are sufficient. In television broadcasting, only one complete side band and the carrier are transmitted for the video, or picture, signal. In many point-to-point transoceanic telephone installations, one side band alone is transmitted, the carrier and remaining side band being suppressed by circuits in the transmitter. The carrier is resupplied at the receiving station by a closely regulated oscillator. Carrier suppression is impractical for broadcast purposes, because the close regulation required for the reintroduced carrier would increase unduly the cost of the receiver. The objects of all side-band- and carrier-suppression schemes are to reduce the band width required for transmission, thus making room for more stations, and to concentrate a greater proportion of the radiated energy in the side-band frequencies, which are the frequencies that convey the message. Single side-band transmission requires a minimum band width, concentrates all the radiated

¹ W. L. EVERITT, "Communication Engineering," 2d ed., pp. 385-387, McGraw-Hill Book Company, Inc., 1937.

power in the message-carrying frequencies, and in addition has some measure of secrecy, since in the ordinary receiver the detector would produce an unintelligible output due to the absence of the carrier.

DETECTION IN MULTIELEMENT TUBES

Triodes, tetrodes, and pentodes may be used as detectors. When the rectification takes place in the grid circuit, the detection is called grid-circuit detection; when the rectification takes place in the plate circuit, the detection is called plate-circuit detection.

19. Grid-circuit Detection.—The circuit most commonly used for grid-circuit detection, Fig. 19.1, is known as a “grid-leak” detector circuit, the resistor in the grid circuit being the “grid leak,” a name acquired in the early days of radio. Essentially, the grid and cathode of the tube act as a diode, developing an audio and a d-c voltage across the RC combination that acts as the diode load, equivalent in all respects to the diode loads of Figs. 4.1 and 13.1. The only difference is that the grid leak is in series at the anode terminal of the rectifying element (in this case the grid of a triode) rather than at the cathode.

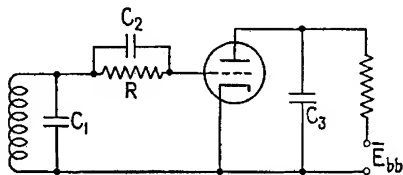


FIG. 19.1.—Grid-leak detector circuit. C_1 is a tuning capacitor; C_2 and C_3 are r-f by-pass capacitors.

Square-law Detection.—When the applied r-f voltage is small, the grid acts as a square-law diode detector. The quiescent grid voltage is slightly negative and is determined by the intersection of the i_c - e_c characteristic and the d-c load line corresponding to the value of the grid-leak resistance, as in Fig. 11.1 for the diode. When an unmodulated signal is impressed on the grid, the average grid current increases because of the square-law rectification. This increased current causes the average grid voltage to shift in the negative direction owing to the increased voltage drop through the external grid leak. The increased negative bias, resulting from rectification in the grid circuit, decreases the plate current. In effect, the d-c voltage developed across the grid leak is amplified and appears usually as a larger voltage across the plate load.

When the signal applied to the detector is amplitude-modulated, the average current through the grid-leak resistance fluctuates in a manner similar to that indicated in Fig. 15.1. If Z_i is the audio-

frequency impedance of the grid-leak-grid-capacitor combination, the rms audio-frequency voltage appearing between the grid and cathode terminals is $I_1 Z_1$, where I_1 is the audio-frequency component of grid current. The problem of determining the voltage appearing at the grid as a result of detection is similar to the problem of determining the voltage across the load of the diode detector. The grid takes the place of the diode, and the grid leak and grid capacitor (R_c and C_c , Fig. 19.1) take the place of the diode load.

Since the effect of the increase in grid current due to rectification is to increase the IR drop across the grid leak, the average grid voltage increases in the negative direction as the signal becomes stronger. The average plate current decreases. When the signal impressed is an amplitude-modulated wave, the decrease in plate current fluctuates at an audio rate producing an amplified audio-frequency voltage across the audio-frequency load in the plate circuit. Owing to the low value of the grid-bias voltage, the amplification is not linear.

There is another secondary effect to be taken into account. Owing to the r-f voltage impressed on the grid, there are r-f components of plate current. If the i_b-e_c characteristics are concave upward, square-law rectification takes place in the plate circuit, tending to increase the average plate current. This tendency is in opposition to the decrease caused by the negative voltage developed in the grid circuit. However, the grid-circuit rectification predominates.

Square-law grid-circuit detection at one time was employed extensively in a circuit known as the regenerative detector in which r-f feedback was used to obtain r-f amplification and grid-circuit detection at the same time. This circuit was characterized by a very high sensitivity and a large amount of a-f distortion. The development of satisfactory r-f amplifiers has removed the necessity in ordinary receivers for detectors of extreme sensitivity, so that the grid-leak weak-signal detector has been displaced by less sensitive but more nearly linear types of detectors, especially the diode.

Grid-leak Power Detector.—If the signal impressed on the grid-leak detector circuit Fig. 19.1, is increased to above 1 volt, say, the grid acts as a large-signal diode detector, developing a voltage across the grid leak that tends to follow the envelope of the applied r-f voltage as in the diode large-signal detector. The average negative grid voltage will be practically proportional to the carrier

amplitude, and the audio-frequency grid voltage (for an amplitude-modulated wave) will be practically proportional to the variations in amplitude as indicated by the envelope of the applied r-f wave. The d-c and audio-frequency voltages developed across the grid leak and grid capacitor can be determined by the methods employed to determine the voltage across the load of a large-signal diode detector.

In the detection of an amplitude-modulated wave the negative d-c voltage developed across the grid leak acts as the bias on the grid, and the audio-frequency voltage developed across the grid leak acts as an audio-frequency signal that is amplified by the tube as in an ordinary audio-frequency amplifier stage. The r-f components of plate current are by-passed by a capacitor connected between plate and cathode, in some cases a filter circuit or an r-f choke coil is interposed between the plate and the a-f load impedance. Over a somewhat restricted range of r-f amplitude, the bias voltage developed across the grid leak will be of the correct value for distortionless audio-frequency amplification, and then the grid-leak power detector will be practically as good as a combination of a diode large-signal detector and one stage of audio-frequency amplification. Outside this restricted range, the grid-bias voltage will be too low or too high for distortionless amplification.

20. Plate-circuit Detection.—Plate-circuit detectors are still used where it is necessary to place a minimum of load on the r-f source. Their operation, as in the previous detector circuits, falls into two general classes, square-law and linear.

Square-law Plate Detectors.—In the square-law plate detector, the biasing voltages on the tube electrodes are chosen so that the i_b - e_c characteristic is parabolic. With an unmodulated r-f signal impressed, the increase in average plate current is proportional to the square of the amplitude of the r-f voltage. This fact is often utilized in the construction of vacuum-tube voltmeters designed to measure rms values.

When the impressed r-f voltage is amplitude-modulated, the square-law rectification yields currents having the frequencies tabulated in Table 15.1, and for a given percentage of modulation the amplitude of the fundamental audio-frequency current is proportional to the square of the carrier voltage. This behavior has been made use of in certain receivers designed for use in connection with radio ranges, to exaggerate the effect of differences in signal

strengths on the output of the receiver and thus to assist the pilot in keeping his ship "on course."

Linear Plate-circuit Detectors, or Biased Detectors.—If a triode or pentode is biased to cutoff and there is no load in series with the plate, half-wave rectification takes place in the plate circuit when an r-f voltage is applied and thus an average plate current flows that is approximately proportional to the amplitude of the r-f signal. The tube may be considered to operate in Class *B*; i.e., plate current flows during one-half the r-f cycle. If a load is connected in series with the plate, there will be an average voltage drop across the load that will cause the average plate potential to decrease. The larger the signal, the greater the decrease. For

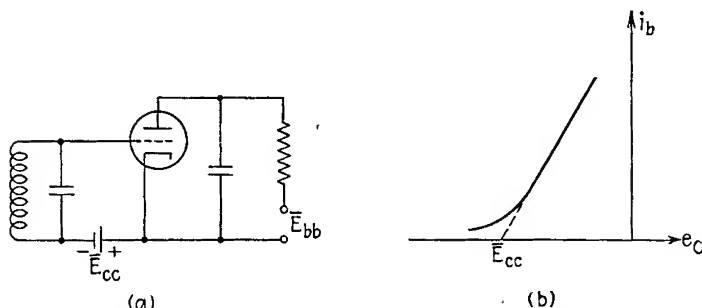


FIG. 20.1.—(a) Biased detector circuit; (b) showing E_{cc} adjusted to "projected" cutoff.

the triode, the decrease in average plate voltage in conjunction with the fixed grid bias causes the plate current to flow for a smaller portion of the r-f cycle, so that the tube operates more or less in Class *C*. If the signal is large and the plate-load resistance is large, the interval in the r-f cycle during which plate current flows may be extremely brief, as in the large-signal diode detector. For a pentode, however, the plate voltage has relatively little effect on the plate current, so that the operation remains Class *B*.

The operation of the linear plate-circuit detector, commonly called the biased detector, may be described by a rectification diagram similar to that for the diode. In practice, the tube is not biased exactly at cutoff because the i_b - e_c characteristics are rounded off at the foot. Instead, the tube is biased to the "projected" cutoff, Fig. 20.1b, or, as recommended in the tube manuals, is biased so that the zero-signal plate current has some stated value. This value is approximately 0.2 ma for triodes and varies for a sharp-cut-off pentode depending upon the plate and screen voltage. Remote-

cutoff pentodes are not satisfactory as biased detectors since their cutoff is not sharply defined.

A rectification diagram for a 6C5 triode operated as a biased detector is shown in Fig. 20.2. The characteristics of Fig. 20.2 are more curved than those of Fig. 6.1, and their spacing indicates that for a given load and r-f signal the triode detector develops a voltage approximately μ times the voltage developed by the diode detector. The curves of Fig. 20.2 are not as equally spaced as those

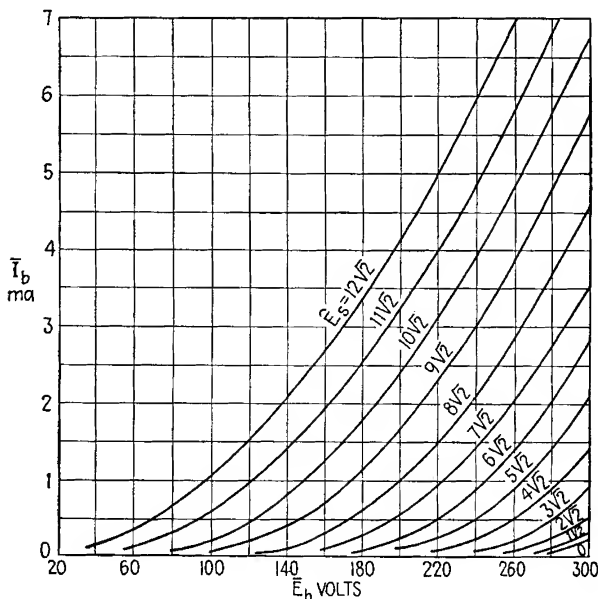


FIG. 20.2.—Rectification diagram for 6C5 biased detector; $\bar{E}_c = -16.5$ volts. of Fig. 6.1, so that the biased detector has more distortion than the diode detector.

A rectification diagram for a 6J7 pentode used as a biased detector is shown in Fig. 20.3. The characteristics are practically horizontal lines because plate current is relatively independent of plate voltage. The pentode biased detector is very sensitive and is capable of developing enough voltage to drive a power pentode¹ without any intervening audio-frequency amplifier stage.

The chief advantages of the plate-circuit large-signal detectors is their sensitivity and their high input impedance (if the grid is

¹ RCA Receiving Tube Manual, Technical Series RC-14, p. 114, RCA Manufacturing Co., Inc., 1940.

not driven positive). Compared with the large-signal diode, they have greater distortion.

Infinite-impedance Detector.—The “infinite-impedance detector,” Fig. 20.4, is essentially a nonlinear cathode follower in which the voltage produced by the rectification process appears across the

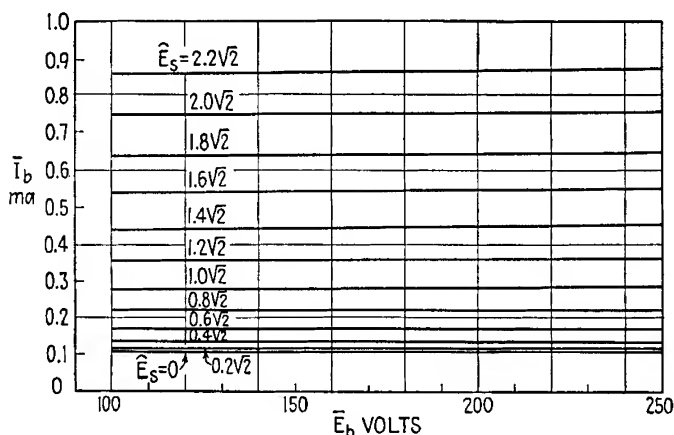


FIG. 20.3.—Rectification diagram for a 6J7 pentode; grid bias = -3 volts, screen voltage = $+50$ volts.

cathode resistor. The name implies that the input impedance is high because the resistance R_k in series with the cathode ensures that the grid is always negative with respect to the cathode, even when the signal is large.

A rectification diagram for a 6J5-G tube as an infinite-imped-

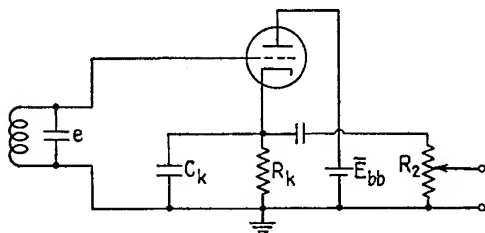


FIG. 20.4.—Infinite-impedance detector with potentiometer for manual volume control.

ance detector is shown in Fig. 20.5. The data for such a diagram may be determined by making C_k very large and measuring the voltage across and the current in R_k as R_k is varied, keeping E_{bb} and the applied r-f voltage constant. The cathode current I_k is plotted against the voltage of “ground” with respect to the cathode,

$\bar{E}_k = -\bar{I}_k R_k$. This diagram thus drawn resembles and is used in the same manner as the rectification diagram for the diode large-signal detector. The load line APB , Fig. 20.5, is drawn for $R_k = 100,000$ ohms and is the counterpart of the line APB , Fig. 6.1. The a-f load line CPD , Fig. 20.5, is the counterpart of the line CPD ,

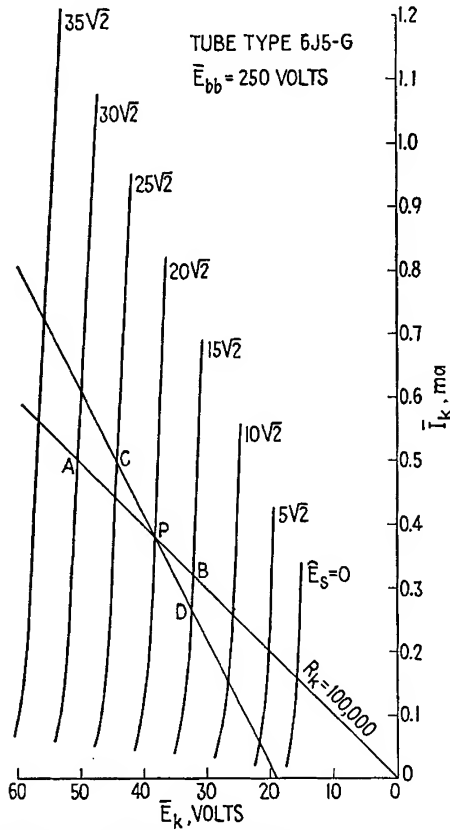


FIG. 20.5.—Rectification diagram for a 6J5-G tube as an infinite-impedance detector.

Fig. 6.1, and corresponds to an a-f load resistance of 50,000 ohms. The output voltage of an infinite impedance detector is approximately the same as that for a diode. The “detection μ ,” or μ_d , for Fig. 20.5 is approximately 0.9, and the r_d is approximately 5,000 ohms.

If the low-frequency impedance of the diode load is not too low compared with the d-c load resistance, the foot of the low-frequency

or, a-f, load line CD , Fig. 20.5, will intersect the $\bar{E}_s = 0$ curve before it intersects the horizontal axis of the diagram. Under these conditions the infinite-impedance detector can detect a 100 per cent modulated wave without clipping. However, if the audio-frequency or video-frequency load impedance is too low, clipping can occur just as in the diode detector. The rectification characteristics are more crowded in the region of zero carrier voltage than at the higher signal levels.

The infinite-impedance detector cannot furnish AVC voltage since the ungrounded side of the cathode resistance R_k has an average positive potential with respect to ground. The cathode terminal of R_k cannot be grounded because the power supply for \bar{E}_{bb} is connected to the other terminal. The chief advantages of the infinite-impedance detector are the avoidance of clipping (however, the output is not free from distortion at high percentages of modulation) and the high input impedance.

DETECTION OF FREQUENCY-MODULATED SIGNALS

The detector in a frequency-modulation receiver is usually a large-signal detector of the diode type. Such detectors respond to amplitude modulation only. This means that the detector must be preceded by a circuit which will impart to the signal a change in amplitude which varies in synchronism with the change in frequency. This circuit is called a discriminator circuit. The discriminator circuit and the rectifying element or elements constitute the detector of frequency modulation.

21. Tuned Circuit as a Discriminator.—A tuned circuit has a different response to different frequencies and can be used as a discriminator, Fig. 21.1. The LC circuit is tuned to a frequency above or below the carrier frequency of the frequency-modulated wave to be detected. As the frequency swings back and forth, the voltage \bar{E}_{AB} , Fig. 21.1, rises and falls, the amount of rise and fall depending upon the frequency swing and the slope of the frequency-response curve of the LC circuit. The r-f voltage across AB then is amplitude-modulated as well as frequency-modulated. The diode detector responds to the amplitude modulation. This circuit is not completely satisfactory since the sides of the resonance curve are not straight; the envelop of the voltage across AB does not vary linearly with the variations in the impressed frequency. This introduces nonlinear distortion into the detector output.

Further, the range over which the frequency may swing without causing excessive distortion is rather limited.

A conventional receiver designed for amplitude-modulated waves may yield an intelligible (but by no means perfect) output in response to a frequency-modulated signal if tuned to one side of the carrier frequency. This is due to the inherent frequency-discriminating properties of the tuned circuits in the receiver.

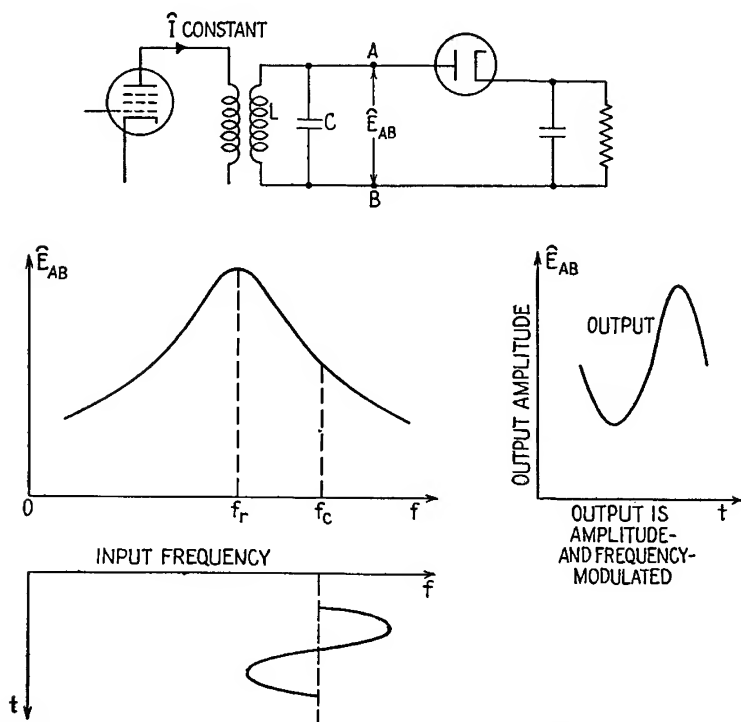


FIG. 21.1.—Showing the introduction of amplitude modulation caused by a tuned circuit in response to a frequency-modulated wave.

22. Double-tuned Center-tapped Discriminator Circuit.—Many discriminator circuits may be designed to have an amplitude response that varies linearly with changes in the impressed frequency.¹ Figure 22.1 shows the type of detector circuit most commonly used in commercial receivers.

¹ A. HUND, "Frequency Modulation," pp. 195–209, McGraw-Hill Book Company, Inc., 1942; J. F. RIDER, "Automatic Frequency Control Systems," pp. 20–52, published by author, 1937; F. E. TERMAN, "Radio Engineers' Handbook," pp. 585–588, 654–656, McGraw-Hill Book Company, Inc., 1943.

The discriminator part of the detector circuit comprises the tuned-primary tuned-secondary transformer together with the capacitor C connecting the ungrounded side of the primary coil to the center tap on the secondary coil. The capacitors C , C_3 , C_4 , C_5

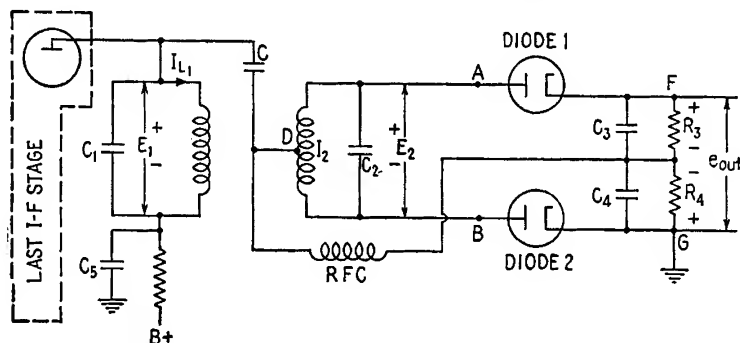


FIG. 22.1.—Double-tuned, center-tapped discriminator circuit.

are large enough to have negligible reactance to the frequency-modulated wave. Both primary and secondary coils are tuned to the carrier frequency.

At the carrier frequency the voltage E_2 is practically in quadrature with the voltage E_1 . The

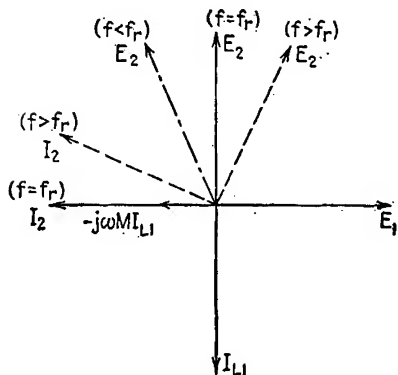


FIG. 22.2.—Phase relationships in the discriminator circuit of Fig. 22.1. Position of E_2 at resonance given by solid line; at a higher frequency, by the dashed line; at a lower frequency, by the dot-dash line.

The over-all result is that E_2 leads E_1 by 90° , Fig. 22.2.

At a frequency higher than the resonant frequency, the current I_{L1} still lags E_1 by 90° (since the primary coil remains inductive).

dissipation being assumed to be small, the inductance of the primary coil causes the current I_{L1} to lag by 90° the voltage E_1 developed across the primary, Fig. 22.2. The emf induced in the secondary coil is $-j\omega MI_{L1}$. The circulating current I_2 in the secondary is in phase with the induced emf since the circuit is in resonance. The external voltage E_2 across the secondary will be the voltage drop across the capacitor C_2 . The diode currents being neglected, E_2 lags by 90° the current through

and the induced emf $-j\omega MI_1$ remains the same in phase. The magnitude may change; the purpose of Fig. 22.2 is to show the phase relationships, the changes in amplitudes being neglected. The secondary circuit presents an inductive reactance to the induced emf since the reactance of the coil has increased in magnitude and the reactance of C_2 has decreased in magnitude. Hence the current I_2 lags the induced emf as shown by the dashed line, Fig. 22.2. The $I_2 X_2$ drop across C_2 is the E_2 indicated by the dashed line, Fig. 22.2. Hence the voltage E_2 now leads E_1 by less than 90° . At a frequency less than the resonant or carrier frequency, E_2 will lead E_1 by more than 90° as shown by the dot-dash line, Fig. 22.2.

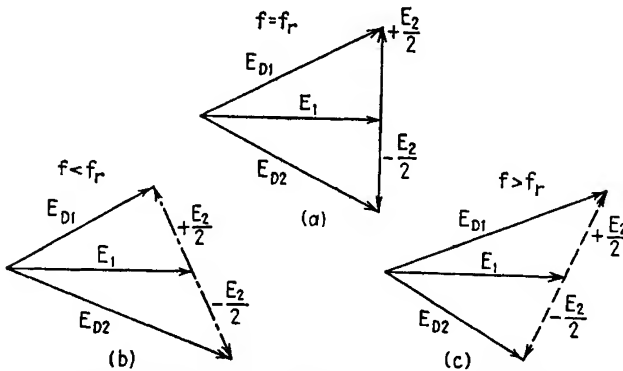


FIG. 22.3.—Showing changes in r-f voltages E_{D1} and E_{D2} impressed on diodes 1 and 2 as the impressed frequency changes.

The r-f voltage impressed on diode 1, Fig. 22.1, is E_1 plus half of E_2 , or E_1 plus the voltage E_{DA} . This addition is effected by the capacitor C , which places the center tap of the secondary at the same r-f potential as the plate terminal of the primary coil. Similarly, the r-f voltage impressed on diode 2 is E_1 plus the voltage E_{DB} , or E_1 minus half of E_2 . The additions are indicated in Fig. 22.3, where the r-f voltages impressed on diodes 1 and 2 are indicated by E_{D1} and E_{D2} .

At resonance, the r-f voltages impressed on diodes 1 and 2 are equal. As a result, the d-c voltages developed across the equal resistors R_3 and R_4 are equal, and the output voltage between F and G , Fig. 22.1 is zero. The radio-frequency choke coil RFC provides the d-c return path for the rectified currents flowing through R_3 and R_4 . R_3 and R_4 are commonly 0.5 to 1 megohm each. The choke coil is replaced sometimes by a resistor of about 50,000 ohms to save the expense of a coil.

When the impressed frequency increases to a value above resonance, the r-f voltage impressed on diode 1 increases, Fig. 22.3*b*, and the rectified voltage across R_3 increases, Fig. 22.1. Simultaneously, the r-f voltage impressed on diode 2 decreases, Fig. 22.3*b*, and the rectified voltage across R_4 , Fig. 22.1, decreases. The net result is an output voltage of positive polarity, point F being positive with respect to ground.

When the impressed frequency decreases to a value below resonance, the voltage impressed on diode 1 decreases and that impressed on diode 2 increases, Fig. 22.3*c*. The rectified voltage across R_3 decreases, that across R_4 increases, and there is a negative output voltage, Fig. 22.1, point F being negative with respect to ground.

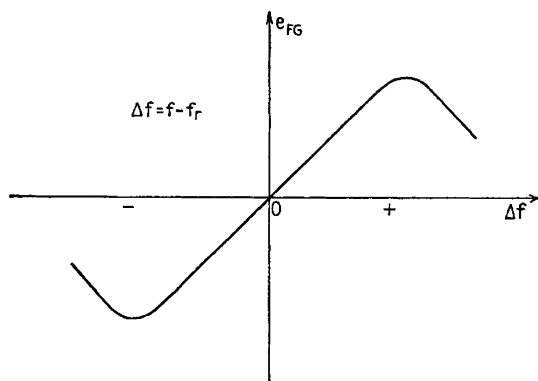


FIG. 22.4.—Relationship between e_{FG} , neglecting r-f ripple, and the frequency deviation Δf for the detector circuit of Fig. 22.1.

As the frequency swings above and below resonance, the potential of point F swings positive and negative with respect to ground G . By suitable design, the potential of point F may be made to vary linearly with variations in the input frequency over a range above and below the resonant frequency, Fig. 22.4. The central portion of the curve of Fig. 22.4 may be given an opposite slope by reversing the polarity of the secondary coil of Fig. 22.1.

Discriminator circuits are used not only in detection but in automatic frequency-control systems as well. The output voltage of the circuit of Fig. 22.1 may be used as a control voltage to actuate a frequency-correcting device whose object is to maintain that frequency to which the frequency-detector circuit has a null response. Such automatic frequency-control systems are employed in both transmitting and receiving apparatus.

CHAPTER XXII

TEST INSTRUMENTS

Test instruments are used mainly to check the operation of equipment and to locate trouble. They are portable, or at least not permanently attached to the equipment, and usually give a quick reading of moderate accuracy. Two general types of measurements are made. One consists of the checking of circuit elements and of fixed voltages and currents, and the other involves following through the circuit a signal that may be self-generated or supplied by an external generator.

1. Direct-current Ammeters.—The moving-coil galvanometer forms the basis for most direct-current measuring devices. On the assumption that it possesses the necessary ruggedness and quick response, the most important characteristic of a galvanometer is its sensitivity, or the deflection per unit current through it. If the terminals of the moving coil are shunted by a low resistance, its current-measuring capacity is increased, as the current through the coil is then a fixed fraction of the total current. Such an instrument is known as an ammeter. It is always placed in series with the conductor through which the current flows. Its resistance must be small in comparison with the rest of the circuit so that its insertion does not materially affect the value of the current.

Although the electric circuits of an ammeter are generally insulated from the case, it is advisable to insert an ammeter at a point as near ground potential as possible in order to reduce the danger of insulation breakdown and injury to the operator. This also reduces the shunting effect of the capacitance of the meter to ground, which is important if high-frequency components are present. Owing to its low resistance an ammeter and the equipment being tested may be seriously damaged if the ammeter is incorrectly placed in a circuit.

2. Direct-current Voltmeters.—A d-c voltmeter consists of a galvanometer in series with a high resistance. The current through it and its deflection are proportional to the impressed voltage. Voltmeters are commonly described by stating the ratio of the

resistance of the instrument to its full-scale reading in volts. This is known as the "ohms/volt" ratio of the instrument, and its reciprocal is the current in amperes required for full-scale deflection. Voltmeters in common use have ohms/volt ratios from 100 to 25,000 corresponding to full-scale currents of from 10 ma to 40 ma.

The smaller the current drain from the source, the smaller the change in voltage caused by the addition of the voltmeter to the circuit. For reasonable accuracy it is necessary that the voltmeter resistance be large in comparison with the internal resistance of the equivalent generator¹ to which it is applied. The internal resistance of most direct-current power supplies is small enough so that voltmeters with relatively low ohms/volt ratings can be used to measure their voltage. The working d-c potentials of tube electrodes, however, are commonly maintained through high-resistance circuits. In measuring such potentials it is important to consider the effect of the added current drain due to the meter in interpreting the reading, even with voltmeters of the highest resistance. The d-c vacuum-tube voltmeter, which has a greatly increased ohms/volt rating, allows many electrode potentials to be measured with reasonable accuracy.

3. Ohmmeters.—The ohmmeter is a most versatile and important test instrument. Its use ranges from simple checks on circuit con-

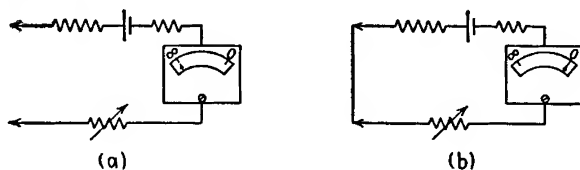


FIG. 3.1.—Series-type ohmmeter circuit.

tinuity to measurements upon inductors and capacitors. It compares an external unknown resistance with an internal standard. The circuit of Fig. 3.1a illustrates one form of the instrument. With the probes at the left open (infinite resistance), the pointer of the instrument is mechanically adjusted to the ∞ mark. With the probes shorted (zero resistance), the full-scale deflection is (electrically) adjusted to the zero mark by means of the variable series resistor, Fig. 3.1b. Any conductor has a resistance between these values and a corresponding intermediate galvanometer deflection. Changes in battery resistance and voltage are compensated to a

¹ A source of potential may be regarded as a generator having an equivalent open-circuit voltage and an equivalent internal resistance.

large extent by the variable-resistance adjustment made with the leads shorted. In other circuits this zero adjustment may be effected by a variable resistance shunted across the meter. The scale markings of the instrument are spaced widely enough to be useful only for a limited range of resistances for any one value of circuit resistance. To cover a wide range of resistances several scales with different values of internal series resistances and galvanometer shunts must be employed. This may be accomplished by a switching arrangement, Fig. 3.2a. The highest resistance that can be measured depends upon the sensitivity of the galvanometer movement. By the use of an external series battery and standard resistor, this range can be extended.

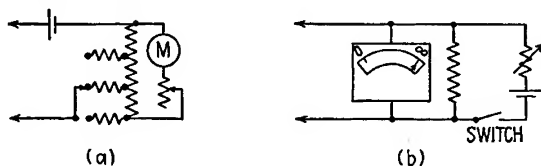


FIG. 3.2.—(a) Multirange and (b) shunt-type ohmmeters.

A circuit that is more satisfactory for the measurement of low resistances places the unknown resistance as a shunt across the indicating galvanometer, Fig. 3.2b. Here the full-scale deflection is adjusted to ∞ with the probes open, and the zero adjustment, if necessary, is made with the probes shorted.

The ohmmeter should never be used on equipment while the power is on. External voltage across the unknown resistor not only makes the reading inaccurate but may ruin the meter. The power should be turned off and any filter capacitors discharged before measurements are made. The leads should never be left shorted in the series-type instrument, and the switch should always be left open in the shunt-type circuit, Fig. 3.2b, to prevent the battery from being run down.

When the ohmmeter of Fig. 3.1a is connected across a capacitor, the ballistic throw of the meter due to the charging current is proportional to the capacitance. This measurement is suitable only for large values of capacitance, but its sensitivity can be increased by using an external series battery. Comparison with the throw obtained from a known capacitor enables the reading to be interpreted.

In checking equipment it is seldom that the magnitude of a capacitor is questioned. The information generally desired is

whether the capacitor is open- or short-circuited or whether its insulation resistance is low. If the capacitance is large enough, the charging kick when an ohmmeter is applied gives an indication that there is no open circuit. The insulation resistance of a paper capacitor can be measured by a sufficiently sensitive ohmmeter and should seldom be lower than 50 megohms. A simple check on the insulation of a paper capacitor is to charge it by application of the ohmmeter probes and then upon quick reversal of the probes note the extent of the ballistic throw of the meter upon discharge. By changing the length of time between charge and discharge a rough indication is obtained of the leakage current through the capacitor.

The leakage resistance of an electrolytic capacitor is much lower than that of a paper capacitor and depends upon the capacitance and voltage rating. The resistance value indicated by an ohmmeter depends upon the voltage of its internal battery and is useful mainly for comparative purposes. The reading obtained depends upon the polarity of the probes. With a multipurpose ohmmeter-voltmeter the polarity marked does not always agree with the actual polarity of the probe voltage. In case of doubt, readings should be made with both probe polarities and the higher resistance value taken as correct.

In tests on an inductor or transformer, the ohmmeter gives immediate information as to whether there is an open- or short-circuit or a low insulation resistance to ground. It may also serve to give an indication of shorted turns if the winding resistance is much lower than its rated value. While resistors vary widely from the value indicated (± 10 per cent is a common tolerance), coils are apt to have a much smaller variation in d-c resistance, owing to their method of manufacture. A center-tapped winding, however, may have unequal resistances on either side of the tap, since the outer part of a coil has a greater length of wire for the same number of turns.

A routine check for trouble may be expedited by a rapid survey of the resistance to ground of several points in the circuit, selected on the basis of their importance and accessibility. Suspected elements then can be examined individually. The circuit diagram must be kept in mind in interpreting ohmmeter readings, with the possibility of parallel circuits always considered. As far as possible, measurements should be made from terminal to terminal to avoid the possibility of overlooking a poorly soldered connection.

4. Tube Testers.—Most common types of tube testers check only the emission of the tube, although provisions are generally made to test for short circuits or low-resistance leakage paths between elements. More expensive instruments give an indication of the control of the grid voltage on the plate current and thus serve to measure something associated with the transconductance of the tube.

Most of the complexity of the instrument is due to the necessity for providing sockets and suitable polarizing voltages (though not necessarily the manufacturers' recommended operating values) for the wide variety of tubes in use. Obviously, such rudimentary measurements can serve mainly to detect a bad tube but will not ensure the correct performance of all the wide variety of functions for which the tube may be used. The final test is its functioning in the circuit, and the simplest check is to replace it by one or more tubes of presumably correct characteristics.

5. A-c Voltmeters and Ammeters.—The a-c voltmeters employed in power work at 60 cps find little use in communication

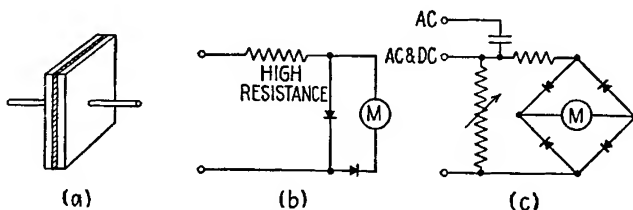


FIG. 5.1.—Rectifier units employed for a-c measurements.

circuits except to measure power-supply and heater voltages. This is due both to their low ohms/volt ratio and the limited range of frequency over which their calibration holds. In their place, rectifier voltmeters are commonly used. These may be designed to have ohms/volt ratios of the order of 1,000 and a constant sensitivity at frequencies up to several thousand cycles/second.

Such instruments employ a copper oxide or similar rectifier unit or units, a series resistor, and a d-c galvanometer movement. The rectifier unit consists of a semiconducting layer formed on a metal backing and another metal plate making contact with the surface of the semiconducting layer, Fig. 5.1a. The contact condition between the metal backing and the semiconducting layer is such that current flows through the unit in one direction more readily than in the other. The thin, partly insulating layer also forms the

dielectric of a capacitor whose capacitance depends upon the area of the plates. At high frequencies this capacitance in effect by-passes the rectifying contact, reducing the rectified direct current and the output reading. By reducing the area of the plates it is possible to increase the frequency range over which the response of the indicating instrument is uniform. Crystal detectors in which the contact area is greatly reduced are useful at very high radio frequencies. Rectifier-type instruments are usable down to very low frequencies. If the voltage to be measured has a d-c component, this must be removed by a series capacitor in the input, generally available at a third input terminal, Fig. 5.1c. When this input is used, the lower frequency limit is determined by the ratio of the impedance of the blocking capacitor to that of the meter circuit.

A rectifier circuit used for measurement purposes is shown in Fig. 5.1b. During the half cycle when the upper terminal is positive, the galvanometer is partly shorted by the conducting parallel rectifier unit so that a negligible current flows through it and its blocking series unit. During the other half cycle the series oxide unit conducts the current through the galvanometer, and the parallel unit is effectively open. Other standard rectifier arrangements are also used. In Fig. 5.1c a full-wave copper-oxide rectifier circuit is combined with a load resistor and becomes an output meter, a device that absorbs and measures the power output of an amplifier or oscillator. The amplifier output is fed into the load resistor within the output meter and the power evaluated in terms of the voltage developed across it. The load resistance frequently is variable to enable the determination of the resistance for best matching conditions. For one particular value of the load (frequently 600 ohms) the scale of the instrument may be calibrated directly in decibels output with respect to a zero-decibel level of arbitrary value.

A cathode-ray oscillograph can be used to give an approximate measure of an a-c voltage. The length of the vertical trace on the screen can be compared with one produced by a known voltage. Its particular advantage is its large input impedance and the high sensitivity obtainable due to its amplifiers. Direct connection to the deflecting plates of a cathode-ray oscillograph allows a simultaneous measurement of d-c and a-c components and extends greatly the useful frequency range.

In measuring alternating currents it is important that the meter introduced in the circuit have a minimum series impedance and a

minimum shunting admittance to ground. This may be attained by introducing a short, straight, thin wire in series with the circuit. This wire is heated by the alternating current, and a fine-wire thermocouple attached to its mid-point generates a direct voltage determined by the heater temperature and therefore by the a-c current. The millivoltmeter that reads the d-c thermocouple voltage may be calibrated in rms units by sending known d-c or low-frequency a-c current through the heater. A simple instrument of this sort combines heater, junction, and millivoltmeter in one container.

Greater accuracy and stability can be obtained if the heater and junction are mounted in a small evacuated glass envelope. This removes the uncertain effects of heat losses due to air and its motion. The temperature of the heater is increased and made more constant for a given heating current. This vacuum thermocouple unit can be enclosed in the millivoltmeter box or mounted in a smaller external unit to allow its insertion directly in the circuit with a minimum length of leads carrying the current to be measured.

When properly calibrated, the thermocouple type of instrument gives the rms value of the current for any waveform, as the heating effect varies with the average of the square of the instantaneous current. It may be used with accuracy at radio frequencies if proper precautions are taken. The reading will be altered by the skin effect due to the change in heater resistance. Therefore, as the frequency range is increased, the diameter of the wire must be decreased. The consequent reduction in current-carrying capacity makes it difficult to measure large currents at radio frequencies by thermocouple methods.

One disadvantage of the vacuum thermocouple unit is its small overload capacity. The wire may be heated in normal use near its safe limit, and relatively small increases above this value may cause the couple to burn out. If such a burnt-out couple is part of a calibrated meter, the whole instrument frequently must be returned to the manufacturer for recalibration. Therefore, it is important to watch the meter carefully in making adjustments, particularly in resonant circuits.

VACUUM-TUBE VOLTMETERS

The combination of an input circuit containing one or more vacuum tubes with a conventional galvanometer movement constitutes a vacuum-tube voltmeter (VTVM), an instrument of

greatly increased input resistance and sensitivity as compared with a simple indicating instrument. A wide variety of such circuits have been evolved for both a-c and d-c measurements,¹ each having special features and limitations. The discussion here is restricted to circuits employed in commonly used test instruments.

6. D-c Measurements.—So long as the input grid of a vacuum tube is held negative, the current drawn by it is extremely small, being due mainly to leakage over the surface of the insulation and the collection of electrons or positive ions by the grid. The electron current is reduced by making the grid sufficiently negative, and the positive-ion current, due to ionization of the residual gas by the electron current to the plate, may be decreased by lowering the

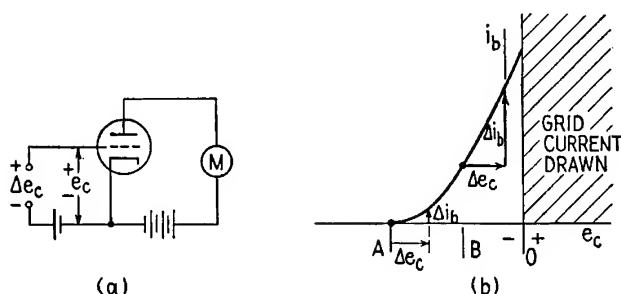


FIG. 6.1.—Simple d-c vacuum-tube voltmeter circuit related to the $i_b = e_c$ -characteristic curve of the tube.

plate voltage. The input grid resistance of conventional tubes is sufficiently high to make possible d-c vacuum-tube voltmeters having ohms/volt ratings of many megohms. Specially designed tubes are available that yield even higher values. The relatively large change of plate current in vacuum tubes for small changes in grid potential (a property measured by the transconductance of the tube) makes possible an instrument of great sensitivity while retaining the ruggedness and quick response of the relatively insensitive galvanometer movement. However, in low-resistance circuits where the unusual characteristics of a vacuum-tube voltmeter are not required, conventional instruments will be found more convenient and in general more accurate.

The simple circuit of Fig. 6.1a serves to illustrate some of the

¹ H. J. REICH, "Theory and Application of Electron Tubes," 2d ed., Chap. XV, McGraw-Hill Book Company, Inc., 1944; F. E. TERMAN, "Radio Engineers' Handbook," p. 929, McGraw-Hill Book Company, Inc., 1943; J. F. RIDER, "Vacuum Tube Voltmeters," published by author, 1941.

problems involved in d-c vacuum-tube voltmeters. If the control grid is biased to cutoff, as at point *A*, Fig. 6.1*b*, a positive increment in grid voltage (Δe_c) produces a corresponding increase of plate current (Δi_b). The resulting voltage scale on the meter would not be linear, however, owing to the curved i_b - e_c characteristic. If the grid is biased at *B* in the linear portion of the characteristic, equal increments in grid voltage would give equal increments in plate current, but they are superimposed on the large quiescent plate current and cannot be measured accurately.

Several circuits make it possible to balance out the steady plate current, so that a more sensitive meter can be used to measure only its changes. In the circuit of Fig. 6.2*a* a meter is connected between a tap on the plate battery and an adjustable contact on the plate resistor. With the grid input shorted it is possible to find a point on the plate resistor whose potential with respect to ground is

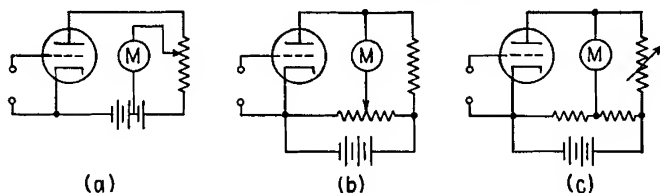


FIG. 6.2.—Various circuits for balancing out the plate current.

the same as that of the battery tap, and no current flows through the meter. If a positive voltage increment now is applied to the grid, most of the corresponding increase in plate current flows through the meter if its resistance is small compared with that of the plate resistor.

The initial zero adjustment with the input terminals shorted is characteristic of many vacuum-tube voltmeters and allows the effect on the calibration of variable conditions to be reduced. In this particular circuit moving the tap on the high-resistance plate load may cause an appreciable change in calibration. This effect is reduced in the circuits of Figs. 6.2*b*, *c*, where a tap is provided on a relatively low resistance shunt across the plate supply. The result is a bridge-type circuit in which the tube is one of the arms, and the off-balance current measures the d-c potential applied at the input to the grid.

The resulting instruments are extremely sensitive, but their input-voltage range is restricted, as the grid should not go positive. The zero point also will be found to drift continually. This results

from the delicate balance accomplished in the zero adjustment and the fact that circuit changes other than those in grid voltage will disturb the balance. Slight changes in temperature, battery voltage, or emission contribute to drifts in the zero.

By the use of negative feedback many of these difficulties can be avoided. In Fig. 6.3a the degeneration introduced in the cathode resistor extends the range of input voltage and makes the grid-voltage-plate-current relation more nearly linear. It also makes the circuit behavior less dependent upon the tube and battery characteristics. In this circuit the meter deflection may be adjusted to a mid-scale zero when the grid input is shorted. Both positive and negative input voltages then may be read without reversal of leads. In Fig. 6.3b a similar circuit balances out the quiescent plate current at zero input voltage. The same circuit is copied in Fig.

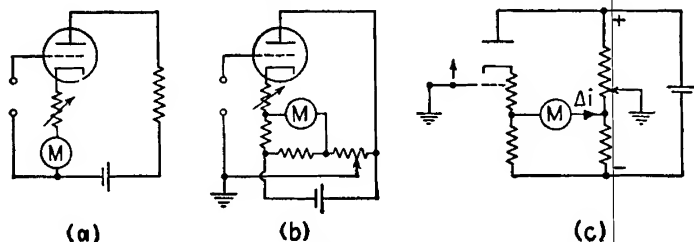


FIG. 6.3.—Vacuum-tube circuits employing feedback.

6.3c in order to show more clearly the potential relations of the various parts. Here each terminal is positioned vertically on the diagram according to its electric potential with respect to ground. The bleeder across the power supply is grounded near the center by means of an adjustable tap, and the control grid at the left is shorted to ground. Under these conditions, adjustment of the tap on the bleeder by the “zero adjust” knob allows the meter current to be made zero. This requires that the left terminal of the meter be at the same potential as its right terminal (same height on the diagram). If the short is removed and the grid is raised in potential, as indicated by the small arrow, the cathode potential also rises slightly and part of this rise is applied to the left terminal of the meter, producing the current shown. It is possible to adjust the sensitivity for purposes of calibration or to obtain various voltage scales by changing the magnitude of the cathode resistor and thus changing the amount of negative feedback. This is an advantage as it avoids the necessity of a voltage divider in the input grid circuit.

Many circuits have been suggested in which the outputs of two tubes are balanced against each other, with the grid of one at fixed potential and the signal applied to the other. Ideally, many of the variables causing drift and instability thus would be canceled out. The major difficulty with such circuits has been in finding tubes with sufficiently similar characteristics. The circuit of Fig. 6.4 avoids this difficulty by supplying sufficient negative feedback to minimize the effect of tube differences. The various electrodes and terminals are located vertically on the diagram in proportion to their electric potential, as in Fig. 6.3c. With the control grid at the left shorted to ground, the circuit has complete symmetry in each branch, and the plates of the two tubes should reach the same

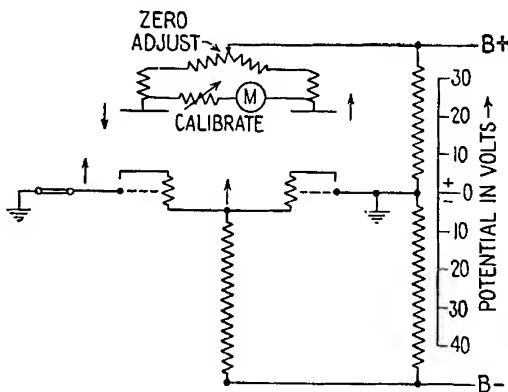


FIG. 6.4.—Diagram of a balanced d-c vacuum-tube voltmeter. Vertical position of elements indicates their quiescent potentials on scale at right.

potential. If current flows in the microammeter between them owing to unequal voltage drops in the plate resistor, it can be adjusted to zero by moving the tap on the potentiometer at the top of the diagram, which thus serves as the “zero adjust” control. If the shorting link on the control grid be removed and a positive voltage increment applied, as shown, the plate current will increase. This causes the plate potential of the left-hand tube to fall and the top of the common cathode resistor to rise in potential. The negative bias on the tube at the right thus is increased, and the potential of its plate rises. Both these changes in plate voltage contribute to a potential difference across the meter and produce a reading that can be calibrated in volts. The calibration can be adjusted by varying the resistance in series with the meter.

7. A-c Measurements.—An a-c vacuum-tube voltmeter could be based upon any circuit that produces a d-c current or voltage

related to the magnitude of the a-c input signal. Circuits employing triodes or pentodes, with their large input grid impedance, would appear well suited for this purpose. Actually, because of the difficulty of providing satisfactory detection over a range of voltages and the problem of providing constant polarizing voltages to maintain calibration in such circuits, the majority of commercial instruments use the simple diode. It is employed in a variety of circuits.

Linear Diode Rectifier Circuit.—One of the simplest vacuum-tube voltmeter circuits is a series combination of a high resistance, a sensitive d-c meter, and a diode, Fig. 7.1a. The operation of this circuit is described in Chap. XXI, Sec. 3. It is referred to there as a linear detector because, for sufficiently large signal voltage and series

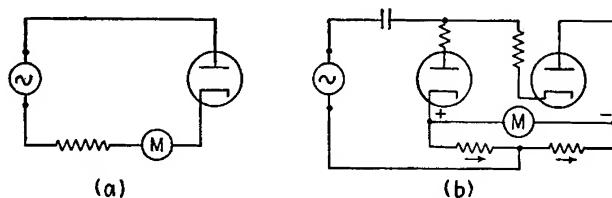


FIG. 7.1.—Diode circuits responding to the average value of the positive loop of sufficiently large signals.

resistance, the component of direct current flowing in the circuit, and therefore the meter deflection, is directly proportional to the area under the positive loop of the applied voltage wave. Under these conditions of operation the current wave closely approximates the form of the positive half of the applied voltage wave except for a constant scale factor. The meter deflection therefore has a response proportional to the average value of the positive loop of the applied signal. The large series resistance makes the action of the circuit approach that of an ideal rectifier by reducing the effects of the curvature of the diode characteristic and of the initial velocity of the electrons. It also reduces the loading effect of the circuit on the voltage source.

For an applied signal of given waveform, generally a sinusoid, the meter may be calibrated to give its magnitude directly. Such calibration is conveniently performed at 60 cps against a conventional a-c meter, and the vacuum-tube voltmeter scale generally is marked in rms units. This 60-cps calibration is accurate for sinusoidal voltages up to frequencies such that the series inductance of the circuit and stray shunting capacitances become effective. Shunt capacitance

across the high resistance causes the reading to increase at high frequencies. A compromise must be reached between linearity and minimum loading on the one hand (which are improved by high resistance) and frequency range on the other (which is difficult to obtain with large series resistance). At high radio frequencies and low values of applied voltage the transit time of the electrons in the diode approaches the period of the signal and causes a reduction in the efficiency of detection. This defect, as well as the shunting capacitance between the plate and cathode of the diode, may be reduced by decreasing the physical dimensions of the diode.

This simple series circuit is seldom used to measure an alternating voltage superimposed on a direct voltage component, owing to difficulty in interpreting the reading. In the full-wave rectifier circuit of Fig. 7.1b a blocking capacitor is inserted, and only the

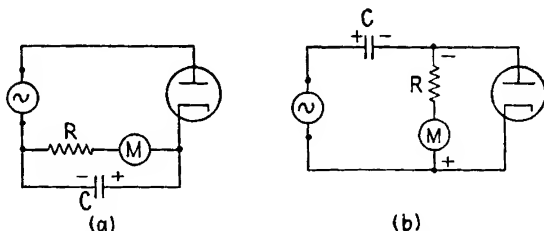


FIG. 7.2.—Series- and shunt-fed diode circuits that, under proper conditions, respond to the peak value of the positive loop of the signal.

a-c component of the voltage acts on the rectifier circuit. The voltage applied to the diode circuit then oscillates about its average value with equal areas above and below the zero axis. During the positive and negative halves of the cycle equal quantities of charge flow through first one and then the other half of the resistance connected across the meter, producing an average d-c voltage across the meter, but no net d-c current in the generator circuit. The voltage developed across the meter is proportional to the average value of either half of the alternating voltage wave.

Series-fed RC Diode Circuit.—If a capacitor is placed across the resistor and meter in Fig. 7.1a, the circuit of Fig. 7.2a is obtained. If the capacitance is sufficiently large, Chap. XXI, Sec. 4, the average potential across the capacitor is only slightly less than the peak value of the positive loop of the applied voltage wave for any waveform. The series combination of the large resistance and the microammeter constitutes a voltmeter that is responsive to the peak value of the applied voltage. The microammeter may be omitted

and the voltage across the RC combination read by a d-c vacuum-tube voltmeter.

If the applied voltage is sinusoidal, the meter reading, which is responsive to the peak or amplitude value, serves to establish the rms value of the voltage and the meter is generally so calibrated. The scale is linear except in the low-voltage range, where the small difference between the average value of the voltage across the capacitor and the peak value of the applied voltage is an appreciable fraction of the voltage amplitude.

The highest frequency for which the calibration holds is limited by much the same factors as in the circuit of Fig. 7.1a except that the stray capacitance across the high resistance is shunted by the relatively large capacitance C and therefore the resistance R may be made quite large. Despite the presence of the by-pass capacitance C the resistance of the circuit is maintained at a value of approximately $R/2$ by the action of the series diode, which conducts over only a small fraction of the period. The low-frequency limit of calibration is set by the value of the time constant RC , which should be large, say one hundred times the period of the lowest frequency that is to be measured accurately. In general, this will be at least as low as 60 cps to enable calibration against standard meters by means of a sinusoidal voltage of this frequency.

It is often desirable in vacuum-tube circuits to measure an alternating component of voltage superimposed on a direct component. As a blocking capacitor cannot be inserted in the circuit of Fig. 7.2a, the following modification is used more commonly in vacuum-tube voltmeters.

Shunt-fed RC Diode Circuit.—In the series-fed circuit the large resistor R slowly discharges the capacitor C in the interval between the short charging pulses that occur when the diode is conducting. If the resistor and meter, Fig. 7.2a, are removed from their position in parallel with the capacitor to one in parallel with the diode, the circuit of Fig. 7.2b is obtained. The operation of the circuit so far as the capacitor is concerned is not altered greatly. It still charges up to the peak of the applied voltage wave and then discharges (chiefly during the nonconducting period of the diode) through R and the generator. If the resistance of the generator is small as compared with R , the time constant of the circuit is unchanged. The direct component of voltage across R again approximates the peak voltage of the positive loop of the applied signal measured from its average value. If a direct component of

voltage is present in the signal, it appears across the blocking capacitor but not across R . The characteristics of the series-fed circuit are maintained therefore except that the direct component of an applied voltage is ignored. In addition to the direct component, an alternating component of voltage now exists across R , which increases its losses somewhat. The input resistance of the circuit on the basis of power considerations is correspondingly decreased from its previous value of $R/2$ to approximately $R/3$.¹

The operation of the circuit is altered in a fundamental manner if the generator resistance R_g is large in comparison with R . If, as is customary, the time constant RC is large in comparison with the period of the impressed voltage, the impedance of C is small in comparison with R and still smaller in comparison with the sum of R_g and R . During the interval when the diode is nonconducting the current through R , and therefore the voltage across it, is a faithful reproduction of the impressed voltage. For even small positive values of voltage across R , and therefore across the diode, the resistance of the diode drops from the very large value corresponding to negative voltages to values that are small in comparison with R . The voltage developed across the combination of R and the diode has an extremely small value during the positive half of the wave as compared with the normal values during the negative half. The effect is to clip the positive half of the wave, approximating the action of an ideal series rectifier except that the voltage across the resistor is a faithful copy of the *negative* portion of the applied voltage wave. Therefore, when the generator resistance is large compared with R , the direct component of voltage across R is proportional to the *average* value of the *negative* loop of the voltage applied to the rectifier circuit, as opposed to the peak value of the positive loop obtained by the same circuit with negligible series resistance in the generator.

It follows that when a shunt-fed RC diode vacuum-tube voltmeter, designed to respond to the peak of an a-c signal, is used to measure the voltage delivered by a source whose internal resistance is of the same order of magnitude as the resistance R in the diode circuit, not only is the voltage read on the vacuum-tube voltmeter reduced by voltage diode action, but the response tends to be proportional to the average value rather than the peak value of the signal.

¹ Chap. XXI, Sec. 4.

8. Characteristics of Typical Commercial Vacuum-tube Voltmeters.—In many commercial vacuum-tube voltmeters the signal is amplified before being rectified and measured. This has the advantage of preserving a high input resistance for the instrument and likewise provides a large diode voltage, which tends to make the meter scale more nearly linear. The gain of the amplifier is generally stabilized by the application of sufficient negative feedback. Such vacuum-tube voltmeters frequently consist of a voltage-divider input unit whose attenuation should be independent of

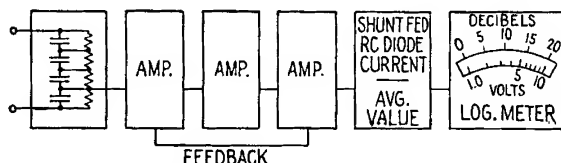


FIG. 8.1.—Ballantine vacuum-tube voltmeter, model 300A.

frequency, followed by one or more stages of amplification feeding into a diode rectifier and meter. Two such instruments will be described. The simplified diagrams are designed to illustrate the operation without regard to circuit details.

The Ballantine model 300A vacuum-tube voltmeter, whose block diagram is shown in Fig. 8.1, has an input attenuator consisting of a resistance-capacitance network whose sections have equal time constants so that the voltage-divider action is independent of frequency over a wide range. Rectification occurs in a shunt-fed

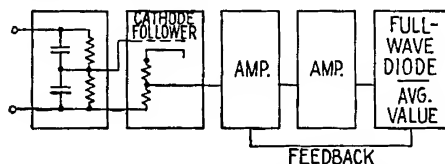


FIG. 8.2.—Hewlett-Packard vacuum-tube voltmeter, model 400A.

diode circuit, the indications of the instrument being approximately proportional to the average of the (positive) half of the applied alternating voltage to the input terminals. The pole pieces of the meter are so shaped as to give a deflection that is approximately proportional to the logarithm of the current through it. This allows a nearly linear decibel scale to be added to the voltage scale.

In the Hewlett-Packard model 400A vacuum-tube voltmeter whose block diagram is shown in Fig. 8.2, the input voltage is applied to a two-step compensated voltage divider that drives a

cathode follower stage having voltage taps on the cathode resistor. As the resistor is of comparatively low resistance, it needs little frequency compensation. The rectification is accomplished in a full-wave diode circuit similar to that of Fig. 7.1*b*. The meter deflections are proportional therefore to the average value resulting from full-wave rectification of the alternating component of the applied voltage.

In making voltage measurements in high-impedance circuits the leads to the vacuum-tube voltmeter may pick up enough stray voltage to affect the reading. Shielding them adds to the input capacitance to ground and produces additional loading at high

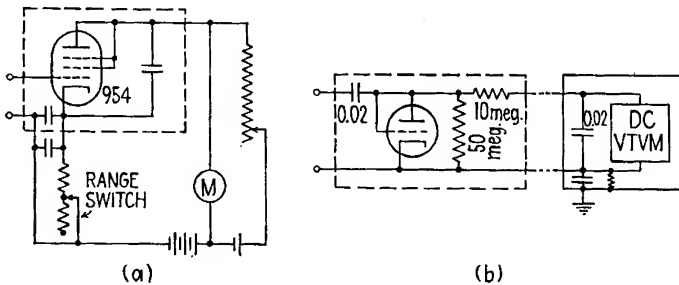


FIG. 8.3.—Probe-type vacuum-tube voltmeters: (a) RCA design; (b) General Radio Company's type 726A. Later models of the General Radio type 726A voltmeter have 60 megohms in shunt and 12 megohms in series.

frequencies. They should be therefore as short as possible. The length of the high-frequency leads can be greatly reduced if the rectifying unit is brought directly in contact with the source terminals. This can be accomplished by the use of a miniature tube, which can be mounted in a probe at the end of a cable that connects the tube to the indicating equipment. The reduced interelectrode capacitances of such a small tube further decreases the shunting capacitance.

The circuit of Fig. 8.3*a* has been suggested by RCA. An acorn pentode connected as a triode operates in a nonlinear region of its (triode) characteristics. The no-signal plate current is balanced out, and the meter reads only the change of d-c plate current due to the applied voltage. The r-f current is confined to the probe (enclosed by dashed lines), and only d-c and low-frequency currents circulate in the other wires, which are collected in a shielded cable going to the power-supply and meter cabinet. Change in range is accomplished by varying the feedback in the cathode circuit. If the signal

source generates a d-c component or does not have d-c continuity, an external blocking capacitor and grid resistor should be added.

An acorn triode connected as a diode is used in the probe circuit of Fig. 8.3*b* employed in the General Radio type 726A vacuum-tube voltmeter. The shunt-fed diode-resistor combination functions as a peak-response rectifier. The 10-megohm resistor in the probe and the capacitor in the control box serve as an r-f filter. The d-c voltage across this capacitor is practically equal to the peak value of the positive loop of the applied a-c voltage. This voltage is read by a one-tube d-c vacuum-tube voltmeter similar to that of Fig. 6.3*b*. The lower input lead is not grounded directly but is connected to ground by a 0.02- μ f capacitor in parallel with a 10-megohm resistor. The case and shielded cable then can be grounded even when the lower input terminal is at a potential other than ground.

9. Interpretation of Vacuum-tube Voltmeter Reading for Non-sinusoidal Waves.—In the vacuum-tube voltmeters described the meter deflection in general is proportional to the average or peak value of either the positive or negative loop of the applied wave. Under some conditions of operation or for some circuits (for example, that of Fig. 8.3*a*) the meter response is not related in so simple a fashion to the waveform. Indeed, the relation may be too complicated to describe usefully and may vary with the amplitude of the impressed voltage. Regardless of the characteristic of the wave to which the meter responds, the scales of most vacuum-tube voltmeters are calibrated in rms volts by the application of pure sinusoidal waves of known magnitudes. For such instruments it is immaterial what property of the wave excites the indicating meter so long as the voltage applied to the vacuum-tube voltmeter is sinusoidal and its frequency lies within the proper range.

If the waveform is not sinusoidal, the readings in general do not give the true rms value of the wave and their proper interpretation requires not only a knowledge of the waveform but also the exact specification of how the meter reading depends upon the characteristics of the wave. Such nonsinusoidal waves are of increasing importance in modern electronic circuits. The departure from a sinusoid may be small owing to the presence of a small harmonic content, or it may be great as in saw-tooth or square waves or repeated pulses of various shapes. If it is desired to obtain the true rms value of such a waveform in order to determine, say, the heating effect in a given resistance, it is necessary to calculate the

relation between the true rms value and the scale reading based on a sinusoidal calibration.

For instruments whose response is accurately proportional to the average or peak value of an applied wave, it is possible to calculate a correction factor K by which the scale reading must be multiplied to obtain the true rms value of a nonsinusoidal voltage.

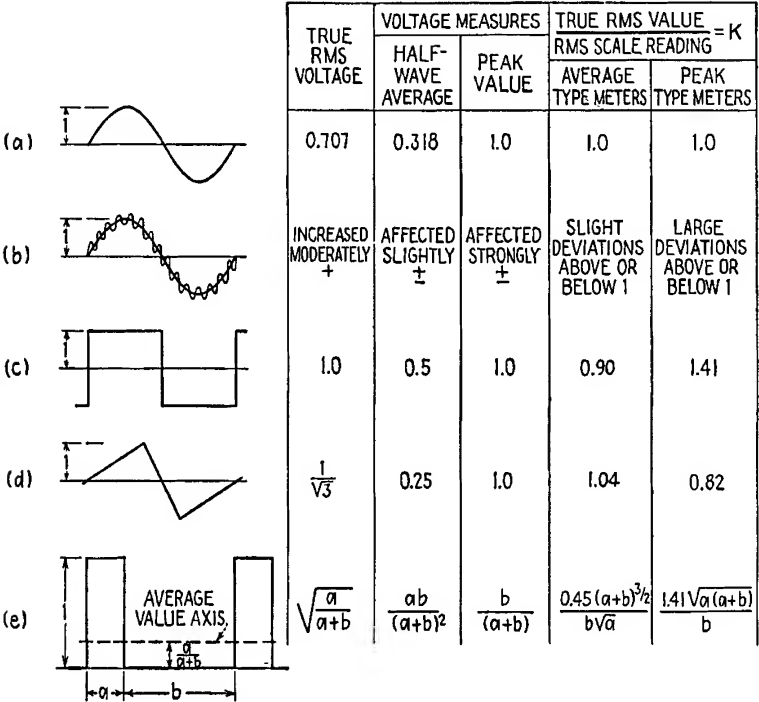


FIG. 9.1.—Various waveforms and their correction factors for average-value and peak-value voltmeters.

The process of calibration of the vacuum-tube voltmeter in rms units against a given sinusoidal voltage may be thought of as equivalent to multiplying the direct-voltage scale of the d-c meter by a calibration factor equal to the true rms value of the sinusoid divided by the direct voltage that it develops in the vacuum-tube-voltmeter circuit. The meter could be calibrated as well directly in terms of a nonsinusoid of known magnitude if it were desired to measure only rms values of the nonsinusoid in question. In this case the calibration factor as before is equal to the rms value of the nonsinusoid divided by the direct voltage it produces in the

circuit. However, when the *nonsinusoid* is applied to an instrument calibrated in terms of a *sinusoid*, the correction factor K is the ratio of the calibration factor of the nonsinusoid to the calibration factor of the sinusoid.

The correction factor K is listed for several waveforms in Fig. 9.1. The true rms voltage is shown for each waveform, together with the corresponding voltage measured by average and peak-responding meters.

For the sinusoid of Fig. 9.1a the correction factor K is obviously unity for an instrument calibrated in terms of a sinusoid. If the waveform contains harmonics as indicated in Fig. 9.1b, the average value may be either increased or decreased by a percentage that is small in comparison with the per cent harmonic content. The peak value will be more strongly affected and may be increased by a per cent equal to the per cent harmonic content if the phase of the harmonic is such that a harmonic peak occurs at the fundamental peak. If the harmonic peak is opposed to the fundamental peak, the reading may be reduced. The true rms value of the wave is increased by the presence of the harmonic but not in the same proportion as the scale reading. In general, it can be said that an average-value instrument is relatively insensitive to the presence of harmonics, while the per cent error of a peak-responsive instrument may be equal to the per cent of the harmonic present. The phase of the harmonic is of importance in determining its effect; as the phase is seldom known, a more exact statement is not called for.

For the square wave of Fig. 9.1c the value of K for an average- and for a peak-value instrument is obtained as the ratio of the calibration factor for a square wave to that for a sinusoid for each instrument.

For an average-value instrument,

$$K = \frac{1/0.5}{0.707/0.318} = 0.90$$

For a peak-value instrument,

$$K = \frac{1/1}{0.707/1.0} = 1.41$$

The values for the triangular wave, Fig. 9.1d, can be checked in similar fashion.

A repeated rectangular pulse, Fig. 9.1e, introduces several new

features. Its d-c component is removed by the blocking capacitors in the input of most a-c vacuum-tube voltmeters, and the voltage applied to the rectifier has the same waveform but oscillates around the average-value axis shown as a dashed line. This average value is located so as to make the area above and below it the same. The wave thus produced has equal average values for the positive and negative parts, but the peak values and the duration of the two loops differ greatly. The numerical values in Fig. 9.1e are based on the assumption that the instrument reading is sensitive only to the positive portion of the loop. Whether this is or is not true depends upon the nature of the detector circuit and the number of stages of amplification. On an average-value instrument with a blocking capacitor the reading would be the same on either peak, or reversing the input leads would not change the deflection. A peak-responding instrument would have its reading affected by reversal of the leads. This is known as turnover. The correction factor of Fig. 9.1e is based on the rms value of the repeated pulse, which is larger than the rms value of the rectangular wave resulting from the removal of the direct component of voltage by the blocking capacitor.

Although this discussion has been concerned with average-value and peak-value vacuum-tube voltmeters it is equally applicable to other instruments that respond to the average or the peak value of the applied voltage. If oxide rectifier units that approach an average-value response are used to measure nonsinusoidal waveforms, the same considerations apply. Some instruments, such as thermocouple meters, dynamometer-type meters, certain special vacuum-tube voltmeters, require no correction, for by their nature they give an rms response for any waveform whose appreciable harmonic content lies within a suitable frequency range.

10. Signal Generators.—In order to test the functioning of many types of radio equipment it is necessary to impress upon them a signal whose characteristics can be adjusted as desired. The nature of the signal depends upon the equipment and the function being studied. In some cases the signal may be a square wave or other nonsinusoidal wave, but more frequently it consists of a sinusoid of variable frequency and amplitude, which may be modulated.

Generators supplying such signals are used to check for faults in a receiver, its alignment, or to measure its over-all response. They may be used as well to determine the characteristics of antennas or other associated equipment. The generators available differ

greatly in their versatility and precision, but all signal generators have as essential parts an oscillator of variable frequency and a means of controlling the magnitude of the output voltage. Provision for modulating the signal is the next most important feature. If measurements are to be made, the voltage output should be known.

The essential features of such an instrument are indicated in the functional diagram of Fig. 10.1. The output attenuator generally consists of a combination of fixed steps and a continuously variable control. If a meter is included, the voltage applied to the stepped attenuator can be adjusted so that accurately known output voltages are available. More complicated signal generators have several oscillators that may be fed into the output, and several types of modulation may be provided. The characteristics of each func-

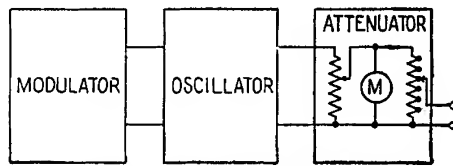


FIG. 10.1.—Functional circuit of a signal generator.

tional unit are discussed below with these possible variations in mind.

Oscillator Characteristics.—The general usefulness of a signal generator is measured in part by the range of fundamental frequencies it covers. Wide range is usually secured by changing coils, each coil determining a band within which the frequency is continuously varied by means of a variable capacitor. Frequently, an additional small variable capacitor is provided by which the frequency can be changed over a small range, thus allowing selectivity measurements to be made with accuracy. The frequency of the oscillator should be constant with time and independent of the load connected to the output. These conditions may be satisfied in part by careful choice of the oscillator circuit and components and by the use of a buffer amplifier stage between the oscillator and the output. Occasionally a quartz crystal is provided to supply fixed frequencies for the calibration of the oscillator or the equipment being tested. When desired, the crystal fundamental and its harmonics yield equally spaced calibration points over a wide range of frequencies.

Modulator Characteristics.—Provision for modulation may be made at one or more fixed frequencies or at an adjustable frequency. This modulation frequency may be generated in a separate internal audio oscillator, which can be fed to the output directly if desired, or provision may be made for modulating with an external signal. If suitable arrangements are made, this external signal may be adjusted over a wide frequency range or may consist of a square wave or a series of pulses. It is desirable that a minimum of frequency modulation should be produced by amplitude modulation of the oscillator, and vice versa. The amount of modulation may be fixed or varied, and in the latter case a meter may be provided to read the per cent modulation.

Frequency modulation may be provided in one or more frequency ranges and with varying band widths depending upon its use. In the conventional amplitude-modulation receiver, it may be used to check the i-f alignment or the operation of automatic frequency-control circuits. For these purposes a narrow-band modulation is satisfactory, but for testing frequency-modulation receivers wide-band frequency modulation is necessary.

Output Characteristics.—The output attenuator may vary from a simple potentiometer to an increasingly complex network as the accuracy desired and frequency range are increased or as the output voltage is reduced to very small values. The calibrating meter may be a thermocouple instrument measuring the current into the attenuator or an a-c vacuum-tube voltmeter that reads the voltage across it. In the latter case the voltmeter is frequently an average-value instrument that is more or less unaffected by the modulation. The output taps may be labeled in volts if the meter is adjusted to a fixed calibration point or may indicate a constant fraction of the voltmeter reading if the input voltage is varied. Where the voltage indicated is the open-circuit voltage across the output terminals, the resistance of the external load connected to the terminals must be large in comparison with the output impedance of the attenuator if the indicated voltage is to be interpreted as the output voltage.

As the frequency is increased and the magnitude of the desired output signal reduced to the order of microvolts, the problem of shielding becomes increasingly important, for it is desired that the only voltage impressed on the circuit being studied be that across the output terminals. The large currents and voltages in the oscillator, unless it is carefully shielded, may induce stray voltages

in the equipment of the same order of magnitude as the terminal potential. This difficulty is avoided by more complete shielding of the oscillator and the use of a shielded cable for the output, it being thus possible for the apparatus under test to be removed some distance from the signal generator. To avoid reflections in the cable at high frequencies it is terminated in a resistor equal to its characteristic resistance.

CHAPTER XXIII

RADIO RECEIVERS

1. General.—A radio or a television receiver selects a desired modulated radio wave and recovers a message from the selected wave. The usefulness of a receiver ordinarily is judged mainly upon four characteristics: *sensitivity*, the minimum input signal required to produce a certain output signal;¹ *selectivity*, the ability of the receiver to discriminate against disturbances at frequencies other than that of the desired signal; *fidelity*, the accuracy with which the receiver reproduces the modulation characteristics of the selected input wave; *signal-to-noise ratio*, the ratio of the signal at a specified point in the receiver to the disturbance at a specified point (usually both being measured at the output). The usefulness of a receiver may be judged by some other qualities, such as radiation, stability, frequency range, AVC action (if any).

The input to a receiver has two constituents: *desired-signal* input and *disturbance* input. The disturbance input consists of electrical effects called atmospheric noise (static) and man-made noise, which are usually nonperiodic, and radiation which has a more or less definite frequency. Likewise, the output has two constituents: *signal* output and disturbance output. The disturbance output consists of amplified disturbance input plus disturbances added by the receiver itself. The signal-to-noise ratio at the output may be less than at the input. On the other hand, certain receivers are relatively insensitive to certain types of disturbance input so that the signal-to-noise ratio actually may be improved in the passage of a signal through the receiver.

Classification of receivers is difficult since there are many different bases of comparison. Four receivers representing basic types are described in the following four sections. Special types, subsidiary parts, and general topics are discussed in the remainder of the chapter.

TYPES NOT EMPLOYING FREQUENCY CONVERSION

2. Regenerative Detector Receiver.—An example of a single-tube regenerative grid-leak detector receiver is shown in Fig. 2.1a.

¹ I.R.E. Standards on Radio Receivers, 1938.

The secondary circuit L_2C_2 is tuned to the carrier frequency of the desired station by adjustment of C_2 . C_1 is an antenna trimmer capacitor. The functions of the grid leak R_1 , grid capacitor C_3 , telephone by-pass capacitor C_4 are discussed in Chap. XXI, Detection. Positive feedback, or *regeneration*, is provided by a coil L_3 , known as a "tickler." The amount of feedback is adjustable. In the receiver shown, the feedback is controlled by varying the coupling between the tickler and secondary coils. Under normal conditions the circuit oscillates as a result of regeneration if the tickler coupling is made sufficiently large. For phone reception the regeneration is kept just under and for CW¹ reception just above

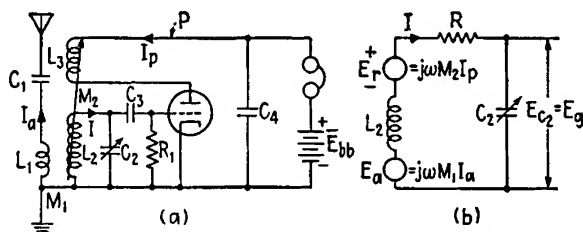


FIG. 2.1.—(a) Regenerative detector circuit and (b) equivalent circuit for the r-f portion of (a).

the point of oscillation. If the detector oscillates during phone reception, a whistle usually is heard, the whistle being a beat note of frequency equal to the difference between the frequencies of the oscillation and of the incoming carrier. The signal then heard in the telephones usually is weak and distorted. However, oscillation is desired in CW reception because the beat note renders the CW signal audible.

An emf $E_a = j\omega M_1 I_a$, Fig. 2.1b, is induced in the secondary circuit by the antenna system and is assisted by an emf $E_r = j\omega M_2 I_p$ induced by regeneration. The secondary current is therefore

$$I = \frac{E_a + E_r}{R} \quad (2.1)$$

where R represents the total dissipative properties of the tuned circuit. The high-frequency voltage developed across the circuit and the strength of the signal heard in the telephones are increased by regeneration.

Alternatively, as in the discussion of oscillators, it may be

¹ CW means continuous-wave telegraphy; during a dot or a dash the wave has constant amplitude.

assumed that a negative resistance R_r is introduced in the secondary circuit through regeneration. On this basis,

$$I = \frac{E_a}{R - R_r} \quad (2.2)$$

and oscillation is understood to begin when R_r becomes equal to R . The negative resistance increases the Q of the secondary circuit, since

$$Q = \frac{\omega L}{R - R_r} \quad (2.3)$$

and therefore the selectivity is improved by regeneration. If the sharpness of the response curve becomes too pronounced, side-band cutting may occur.

In phone reception, the signal is strongest when the circuit is very close to the point of oscillation. However, it is not feasible to operate the detector with the regeneration infinitesimally close to the point of oscillation since a slight change in operating conditions, such as a small rise in plate voltage or a transient disturbance, may cause the circuit to go into oscillation, and the circuit usually remains in oscillation after such an initiating change has disappeared. When the receiver is tuned to a new frequency, the regeneration control may need to be reset because the regeneration and the conditions for oscillation are altered.

In a well-designed regenerative detector receiver, the regeneration control produces a very gradual change in the regeneration and has negligible effect upon the tuning. The regeneration may be controlled by a variety of methods, Fig. 2.2. In modern detectors of this type a tetrode or pentode tube usually is employed, and the regeneration then may be controlled by variation in screen-grid voltage. In some cases the antenna system is also tuned to the frequency of the desired station, and the coupling between antenna and secondary circuits is adjustable. Maximum signal strength and maximum selectivity are obtained by proper adjustment of the coupling. The detector may be followed by one or two stages of audio amplification so as to operate a loudspeaker.

One serious fault of this type of receiver is that it radiates strongly when oscillating, causing "squeals" in near-by receivers. The regenerative detector receiver was used widely in the early days of radio and was standard for several decades in marine radio. When skillfully operated, this type of receiver has good sensitivity,

fair selectivity, reasonable fidelity, and a satisfactory signal-to-noise ratio. It is simple and requires little power for operation. It is well adapted for long- and medium-wave reception and may be used down to a wavelength of the order of 10 m.

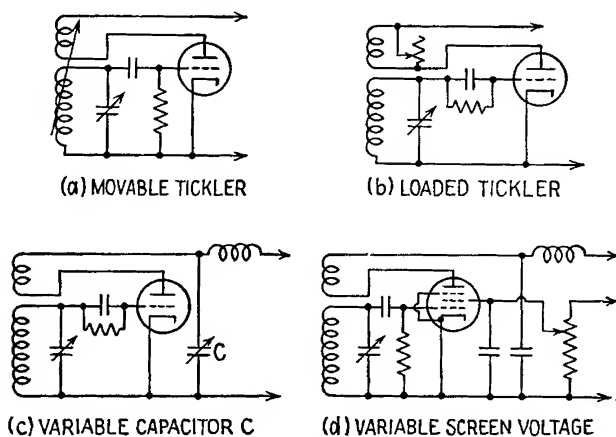


FIG. 2.2.—Methods for control of regeneration.

3. Superregenerative Detector Receiver.—The circuit of Fig. 2.1a may be operated as a *superregenerative* detector by causing the circuit to be alternately oscillatory and nonoscillatory. The operation is different from that of the ordinary regenerative detector, and the sensitivity of the circuit as a detector is enormously greater.

Superregenerative detection may be secured by greatly increasing the coupling and introducing a periodic voltage called a *quench* voltage at the point *P*, Fig. 2.1a, so that the total plate-supply voltage is varied at the quench-frequency rate. The quench frequency f_q is chosen to be above the audible range. The steady plate-supply potential \bar{E}_{bb} and the quench-supply voltage are adjusted so that the tube will oscillate only when the quench voltage reinforces the steady voltage \bar{E}_{bb} ; the supply voltage thus provides conditions that favor oscillation at intervals that occur f_q times per second. In the absence of an incoming signal, a random emf due to thermal agitation, shot effect, incoming disturbances, etc., starts the oscillations at the beginning of each interval favorable to oscillation. Shortly after the onset of oscillation, the plate-supply voltage falls, and the oscillations die away. A sequence of wave-trains of high-frequency oscillations thus occurs at the quench-frequency rate.

Through grid-circuit rectification, a voltage is developed across C_3 , Fig. 2.1a, and the average grid potential is negative during each group of oscillations, tending to produce dips in the average plate current at quench frequency. When the onset of oscillation is controlled by random voltages, the amplitude and duration of the successive wavetrains vary in a random manner because the initiating emf's vary in amplitude and occur at varying intervals after the uniformly spaced moments when conditions become favorable for oscillation. Consequently, the dips in plate current are not uniform, and the plate-supply current varies irregularly. A hissing or frying sound is heard in the headphones, which is characteristic of the superregenerative receiver when it is not tuned to a station.

When an unmodulated high-frequency voltage greater than the random voltages exists across the secondary circuit, this voltage assumes control and initiates the oscillations each time that conditions become favorable. The starting point of each wavetrain becomes locked in with the unmodulated wave. The successive wavetrains become practically uniform, so that the plate-supply current becomes practically steady. Therefore the characteristic hiss disappears when the receiver is tuned to a station whose signal is not too weak.

If the high-frequency wave that controls the detector is a modulated wave, the current passing through the headphones varies in accordance with the modulation. The fidelity depends greatly upon the adjustment of the various factors that control the superregenerative action.

The quench voltage, or the quenching effect, may be obtained within the detector itself; the detector is then said to be *self-quenched*. If the quench voltage is introduced from a separate source, the detector is said to be *separately quenched*. Superregenerative detection may be obtained by employing any condition that results in quench-frequency variation of the transconductance g_m between two limits such that the circuit oscillates at an r-f rate at quench-frequency intervals and energy in the tuned circuit due to one wavetrain is dissipated before the next wavetrain begins.

A receiver containing a separately quenched detector and one stage of audio amplification is shown in Fig. 3.1. The circuit resembles an oscillator circuit, the coupling between plate and grid being governed by the position of the tap on coil L_2 . Tube T_2 and the elements associated with it constitute a quench oscillator.

Quench voltage is developed across the circuit L_4C_4 . The quench voltage and the steady direct voltage obtained from R_4 constitute the plate-supply voltage for the detector circuit. R_2 is an audio-frequency plate-load resistor for the detector circuit. RFC_1 has sufficient reactance to block the high-frequency currents but not the quench-frequency currents. RFC_2 blocks the quench-frequency currents. Capacitor C_6 blocks the direct part of the detector plate voltage from the grid of the audio amplifier. The audio-frequency voltage across R_6 is amplified by tube T_3 .

A separately quenched detector may be operated under any one of several possible conditions called modes.¹ The mode of operation depends upon the frequency, amplitude, and waveform of the

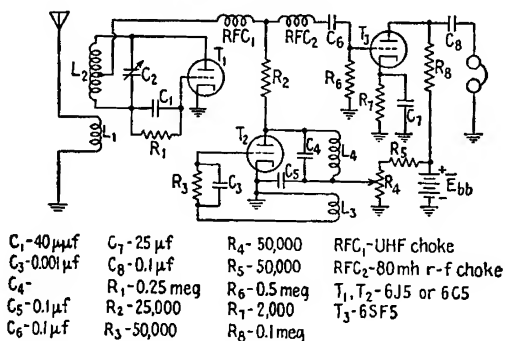


FIG. 3.1.—Separately quenched superregenerative receiver.

quench voltage, the characteristics of the tube and circuit, and the amplitude of the incoming signal. In all cases the quench frequency must be above the audible range but low enough for a considerable number of high-frequency oscillations to take place in each wavetrain. In the *logarithmic* mode, the amplitude of each wavetrain reaches a more or less constant value before decay sets in. In this mode of operation a strong station does not produce much greater detector output than a weak one. The receiver behaves as if it were equipped with AVC, and the AVC action is more intensive and rapid than in the average superheterodyne. On the other hand, the nonlinearity responsible for the AVC effect causes the fidelity of the detector to be far from good.

In the *linear* mode of separately quenched operation, quenching occurs before the growth of oscillation has become limited by

¹ F. W. FRINK, Basic Principles of Superregenerative Reception, *Proc. I.R.E.*, **26**, 76, January, 1938; F. E. TERMAN, "Radio Engineers' Handbook," p. 662, McGraw-Hill Book Company, Inc., 1943.

saturation effects. The output of the detector is related more linearly to the strength of the input signal. The fidelity is better than in the logarithmic mode, but the AVC action is less pronounced.

The circuit of Fig. 3.1, with few alterations, may be made to serve as a self-quenched detector. The quench oscillator is turned off, so that only the direct voltage from R_4 reaches R_2 , and a capacitor large enough to by-pass to ground the quench-frequency voltage to be produced by the detector is attached at the point where R_2 joins the two chokes. The time constant of R_1C_1 then is increased, primarily by raising the value of R_1 , until the circuit operates as an intermittent, or blocking, oscillator, as described in Chap. XV, Sec. 5. The tube oscillates intermittently because the voltage developed across C_1 as a result of grid-circuit rectification periodically "blocks" the tube, *i.e.*, reduces the plate current to zero or to a very low value. The circuit is adjusted until the tube remains blocked for periods long enough for the oscillations to die away completely before the tube becomes unblocked. The length of the blocked period is determined largely by the time constant of R_1C_1 . The action of the self-quenched detector is largely the same as that of the separately quenched detector but is more complicated. For example, the quench frequency becomes a function of the initiating emf and generally increases with the amplitude of the initiating voltage. The amplitude and duration of each wavetrain of oscillations vary when the controlling wave is modulated. The average plate current varies, and the modulation characteristics are thus detected.

The self-quenching effect also may be obtained by providing the detector tube with two tuned circuits so that the tube can oscillate simultaneously at two frequencies, the frequency of the incoming signal and the quench frequency.

The detector tube represents a load on the tuned circuit L_2C_2 , so that the frequency-response curve is broadened and the selectivity is reduced. This load varies during each quench cycle, and for this and other reasons the frequency of oscillation varies slightly during each group of oscillations. Because of the rise and fall in their amplitude, the groups of r-f oscillations have a broad energy spectrum, Chap. XIX, Sec. 13. The net result is a broad radiation spectrum that may have pronounced peaks. Radiation may be reduced or eliminated by use of a preamplifier before the detector. The preamplifier isolates the detector from the antenna system and also improves the selectivity of the receiver.

The superregenerative detector is not satisfactory for CW code reception since it does not produce a beat note. The locally generated oscillations are broken into discrete wavetrains without any definite phase relation between them, so that the necessary periodic conditions for a beat tone are absent. The freedom from beat tones or whistles, the presence and disappearance of the hiss, and the increased sensitivity of the detector are indications of superregenerative action.

The superregenerative detector is relatively insensitive to disturbances of the impulse variety, such as emanate from automobile ignition systems. Apparently a disturbance that occurs at a time other than at the start of a wavetrain produces little effect upon the action of detection. Therefore the signal-to-noise ratio of the superregenerative detector is especially good in applications where there is considerable impulse disturbance.

The greatest sphere of usefulness of the superregenerative detector is in the short-wave and ultra-short-wave regions, where receivers of more conventional design do not operate as effectively. The superregenerative detector will operate in the broadcast band, but in this band other types of receiver are more satisfactory.

4. Tuned-radio-frequency Receivers.—The conventional tuned-radio-frequency (TRF) receiver includes one or more stages of high-frequency amplification before a nonregenerative detector, all of these stages being tuned to the frequency of the incoming signal. Generally, the high-frequency transformers are single-tuned (primary untuned, secondary tuned), though in some cases double tuning has been employed to secure a band-pass response characteristic.

An example of TRF receiver is shown in Fig. 4.1. The first two tubes are r-f pentodes of the remote-cutoff ("supercontrol") type. The following two are r-f pentodes of the sharp-cutoff type, the first serving as detector and the second as a-f voltage amplifier. The last tube is a power-amplifier pentode. The three secondary tuning capacitors are "ganged" on one shaft. The detector is of the large-signal plate-circuit type. Resistance coupling is employed throughout the audio system. The volume is controlled by use of a potentiometer P that regulates the bias voltage on the r-f amplifiers. The RC filters R_1C_1 and R_2C_2 decouple the grid circuits of the r-f amplifiers, C_1 and C_2 serving also as r-f connections between the cathodes and the coils of the tuned circuits. The audio-fre-

quency tubes are self-biased. Decoupling filters are employed in their plate circuits to prevent regeneration and "motorboating." Following the detector, a low-pass filter C_5LC_6 is employed to prevent the radio-frequency signal from passing into the audio-frequency amplifier.

The capacitor C_{11} , which in some cases has a resistor in series with it, improves the matching between the loudspeaker and the output tube and also reduces the high-frequency part of the audio output. The higher tones and higher frequency noise content of the audio output may be reduced to an adjustable level by the use of a variable resistor in series with C_{11} . The higher tones, in themselves, usually are not undesirable. The reproduction at the higher

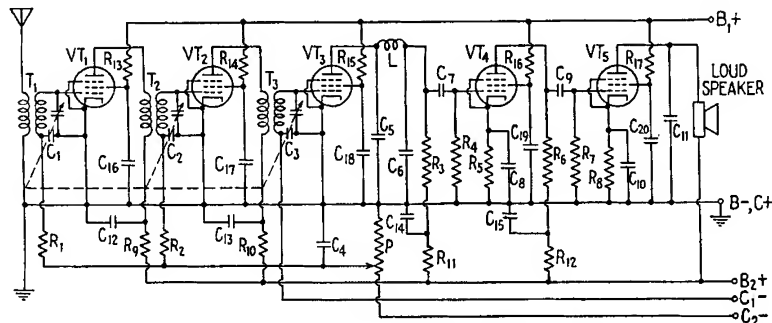


FIG. 4.1.—Tuned-radio-frequency (TRF) receiver.

frequencies may be distorted, or there may be considerable noise at the higher frequencies, and the listener may elect to sacrifice the high audio frequencies in order to eliminate the distortion or noise. However, there is also a personal element in the setting of the tone control.

The TRF receiver can be made to serve as a *communications* receiver, *i.e.*, for CW code reception, by the addition of a so-called "CW-oscillator" circuit. The latter, in this case, is an oscillator whose frequency is controlled by a tuning capacitor ganged with the other tuning capacitors, the frequency of oscillation remaining slightly above or below the frequency to which the receiver is tuned. The output of the CW oscillator is mixed with the incoming signal, usually in the detector circuit, so that a beat tone is produced when an r-f signal is present. The r-f amplifier stages are an effective aid in isolating the oscillations from the antenna system where such isolation is desired.

SUPERHETERODYNE RECEIVERS

5. The Superheterodyne Principle.—A superheterodyne, like any other type of receiver, consists of three parts, high-frequency amplifier, detector, and low-frequency amplifier. Basically, the detector and low-frequency amplifier do not differ from those previously described. The fundamental feature of the superheterodyne is a change of the carrier frequency (frequency conversion) in the high-frequency part of the receiver.

A block diagram of a superheterodyne is shown in Fig. 5.1a. In the *mixer* stage, the incoming signal is nonlinearly combined with the output of a *local oscillator*, and the mixer delivers to the

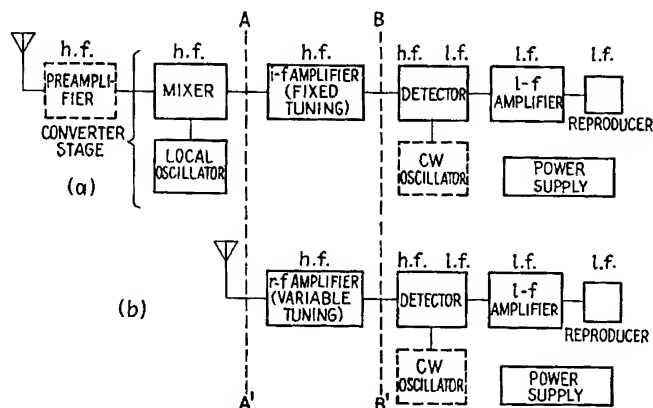


FIG. 5.1.—Block diagram of (a) superheterodyne and (b) tuned-radio-frequency receiver.

i-f amplifier a wave having a new, fixed carrier frequency and the modulation characteristics of the selected incoming signal. The mixer and local oscillator stages together constitute the *converter* stage. The *i-f amplifier* is a high-frequency amplifier having one or more stages and is permanently tuned to the carrier frequency of the converter output, known as the *intermediate frequency*.

An example of a communication-type superheterodyne is shown in Fig. 5.2. The preamplifier and *i-f amplifier* tubes VT_1 and VT_4 are remote-cutoff ("supercontrol") r-f pentodes, while VT_2 is a pentagrid mixer tube. These three tubes are supplied with AVC voltage from an AVC line, a decoupling filter such as R_1C_2 being employed with each tube to prevent interaction among them. VT_3 serves in a tuned-grid oscillator circuit as the local oscillator, for frequency conversion. The three variable capacitors C_3 , C_5 ,

C_{30} are ganged so that the corresponding circuits are tuned simultaneously. The circuits L_1C_3 and L_2C_8 are tuned to the same frequency, that of the desired incoming signal, while $L_3C_{28}C_{30}$ is tuned to a frequency always a fixed amount above or below that of the incoming signal. The difference-frequency output of the converter stage is the input to the i-f amplifier. The i-f input transformer circuits L_4C_{11} and L_5C_{13} are permanently tuned to the intermediate frequency. The i-f wave is amplified by VT_4 and through the permanently tuned i-f output transformer is applied to the detector. The detector in this particular example is a half-wave diode circuit

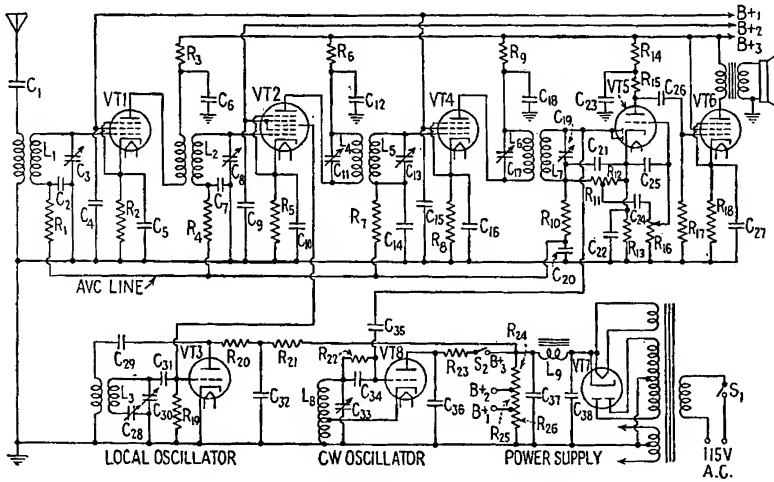


FIG. 5.2.—Communication-type superheterodyne receiver.

that provides AVC voltage in addition to detecting the signal. The triode section of VT_5 and the pentode VT_6 are audio-frequency amplifiers. The power needed by the tubes is obtained from the power supply, in which VT_7 is a full-wave rectifier. Tube VT_8 serves in a CW-oscillator circuit, as described in Sec. 4, except that in this case its output "beats" with the fixed-frequency i-f wave rather than with the incoming r-f wave. Therefore it is not necessary to vary the tuning of the CW oscillator.¹ A switch (not shown) usually is provided to render the AVC system inoperative (for example, by grounding the AVC line) when the CW oscillator is turned on because the CW-oscillator signal will affect the AVC

¹ However, an adjustment is often provided so that the operator can change or adjust the pitch of the beat note.

voltage, and AVC action during code reception may blur the dots and dashes and impair the signal-to-noise ratio.

It may be seen by comparing the block diagrams in Fig. 5.1 that the superheterodyne and TRF receivers are identical from the line BB' to the reproducer. The parts of the two receivers lying between the lines AA' and BB' differ fundamentally in but one respect: the i-f amplifier is tuned to a *fixed* frequency. In medium- and long-wave superheterodynes the most important fact is that the intermediate frequency is *fixed*. (In some all-wave receivers, the frequency of the i-f carrier may actually be higher than that of the r-f carrier; the term "intermediate," however, is still employed.) In many ultra-short-wave superheterodynes, however, the most important fact is that the intermediate frequency is *low* compared with the frequency of the incoming signal, better amplifier characteristics being obtained at the lower frequency.

An amplifier that operates at a fixed frequency can be aligned and adjusted more effectively than an amplifier with variable tuning. The i-f amplifier performs almost equally well no matter where the receiver dial is set. A superheterodyne receiver may be designed with almost any desired band-pass characteristic, and the width of the pass band may be adjustable. For high-fidelity broadcast reception a symmetrical response curve with a broad and nearly uniform response in the pass band may be secured by suitable design and coupling of the i-f transformers. Extreme selectivity for use in code reception can be secured by means of a quartz-crystal filter in the i-f amplifier. All these features are practical only because the intermediate frequency is fixed.

The sensitivity, selectivity, and fidelity of a superheterodyne can be made better than the corresponding properties of a TRF receiver of comparable size and cost. A superheterodyne containing several h-f amplifier stages is more stable than a TRF receiver containing the same number of h-f stages because those in the superheterodyne do not all operate at the same frequency. Also, for the same number of stages, band switching is easier than in a TRF receiver, for fewer stages need to be switched. Further, since the tuning of the i-f part of the h-f amplifier is not changed during the operation of the receiver, the superheterodyne is more likely to remain in proper adjustment.

The superheterodyne has some disadvantages. Difference-frequency whistles due to the presence of the local oscillator may occur at some settings of the tuning dial, whereas a correctly

operating TRF receiver is free from this defect. There may be serious radiation from the local oscillator in the frequency converter. The converter stage sometimes introduces objectionable noise, impairing the signal-to-noise ratio. The superheterodyne is subject to a type of interference known as image interference, discussed in Sec. 9. These disadvantages, however, are not insurmountable and at the present time the superheterodyne, on the whole, is the most useful and versatile type of receiver. The superheterodyne principle is usable at all radio frequencies and for all types of modulation.

6. Superheterodyne Frequency Conversion.—Frequency conversion is accomplished by a process known as *heterodyning*, which is nonlinear mixing of a selected signal and the output of a local oscillator. In early superheterodynes, a grid-leak detector circuit, Fig. 7.1a, was employed as a mixer, and the mixer therefore came to be known as the “first” detector and the signal-recovering stage as the “second” detector. This nomenclature implies that frequency conversion is detection. However, the process more nearly resembles modulation. Therefore, the terms “converter” and “detector” are employed here in place of first and second detector. The converter stage may employ two separate tubes, called *mixer* and *local oscillator* tubes. If a single tube serves as both oscillator and mixer, it is called a *converter* tube.

The incoming signal voltage e_s and the local oscillator voltage e_o may be combined by means of *circuit coupling* and their sum $e = e_s + e_o$ applied to the mixer. In this case the mixer may be referred to as a *single-input* mixer. Examples are shown in Figs. 7.1a and 14.1. Or, mixing may be accomplished by means of *electron coupling*, the voltages e_s and e_o appearing separately at two grids in the same electron stream. Such a mixer may be called a *double-input* mixer.¹ Examples are shown in Figs. 7.1b, c, d. In either type of circuit, new component frequencies are created. The component whose frequency is equal to the difference between oscillator and signal frequencies is generally selected as the intermediate frequency f_i . Thus,

$$f_i = f_o - f_s \quad \text{or} \quad f_i = f_s - f_o \quad (6.1)$$

depending upon which is higher. In the broadcast band, f_o remains above f_s , while in the higher frequency bands f_o may be below f_s .

¹ H. STOCKMAN, Superheterodyne Converter Terminology, *Electronics*, 16, 144, November, 1943.

The choice is largely a matter of convenience in design. The difference in frequency between f_o and f_s usually is maintained at a constant value by ganging the oscillator and r-f tuning capacitors so that they are all turned by the same knob.

The incoming wave, when modulated, consists of a number of components (side bands and carrier). In the process of frequency conversion, all the components are changed in frequency by the same amount. Therefore the separation in frequency between the carrier

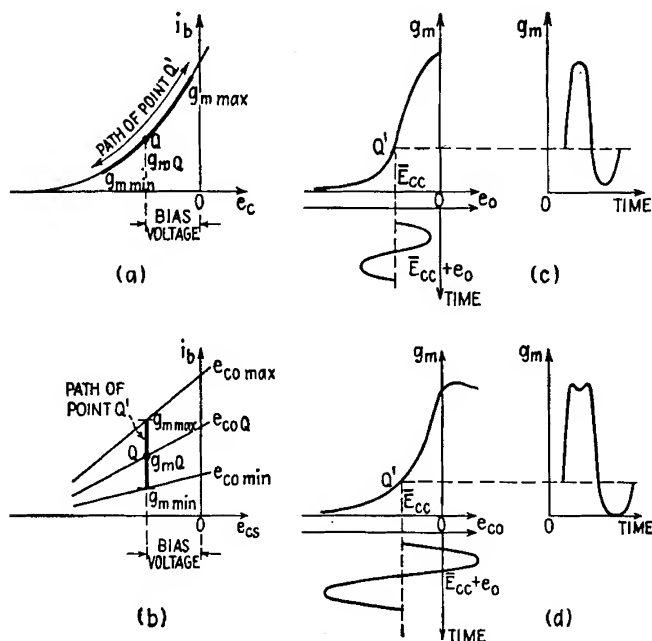


FIG. 6.1.—Operation of frequency converters; (a) and (b): movement of Q point in single- and double-input converters; (c) and (d): variation in value of g_m in single- and double-input converters.

and the side components is not altered. The spectrum of the modulated wave is unchanged except for an equal reduction in the frequency of all the components.

It is customary and convenient to base discussions and analyses of frequency conversion upon periodic variation of the transconductance between the signal grid and plate, caused by the local oscillator voltage. The mixer tube in the single-input mixer of Fig. 7.1a, has a plate-current-grid-voltage characteristic like that in Fig. 6.1a. The point Q is determined by the value of the bias voltage.

The plate potential remains practically constant, for there is practically no signal- or oscillator-frequency voltage across the load (since it is tuned to the frequency f_i) and the voltage developed across the load at the difference frequency is very small. Practically, therefore, the point of operation remains upon the curved path shown. The sum of the bias voltage and e_o determines a moving Q point, designated as Q' , about which the voltage e_s operates with a relatively small amplitude. The magnitude of the changes in plate current produced by the voltage e_s is governed by the slope of the curve, *i.e.*, by the transconductance g_m at the point Q' . The value of g_m varies, Fig. 6.1c, since the i_b - e_c characteristic is curved.

In double-input mixers and converters, the variation in g_m by the local oscillator voltage is achieved in a different manner. Three straight-line portions of more extended characteristics are drawn in Fig. 6.1b to represent the relations between the plate current i_b and the signal-grid voltage e_{cs} for three particular values of oscillator-grid potential e_{co} . The middle characteristic designated by e_{coQ} is the signal-grid i_b - e_c curve upon which the Q point is located when $e_o = 0$. The voltage e_o causes the point Q' to move up and down, for example between the upper and lower characteristics shown in the figure. The value of g_m is given by the slope of the curve upon which the point Q' momentarily is located, and therefore g_m rises and falls in value, Fig. 6.1d, because the curves have unequal slopes. In this case, frequency conversion is not dependent upon curvature in the individual i_b - e_c characteristics.

The principle of conversion in single- and double-input mixers is essentially the same, the circuits differing only in the methods of producing the variation in g_m . For simplicity in analysis, assume that g_m varies linearly with the oscillator voltage, so that

$$g_m = g_{mQ} + k\hat{E}_o \sin \omega_o t \quad (6.2)$$

where g_{mQ} is the Q -point value of g_m and $k\hat{E}_o$ is the amplitude of the variation g_m . The variational part of i_b that is of interest in frequency conversion may be expressed under the previously made assumptions as

$$i = g_m \hat{E}_s \sin \omega_s t \quad (6.3)$$

Combining (6.2) and (6.3),

$$\begin{aligned} i &= (g_{mQ} + k\hat{E}_o \sin \omega_o t) \hat{E}_s \sin \omega_s t \\ &= g_{mQ} \hat{E}_s \sin \omega_s t + k\hat{E}_o \hat{E}_s \sin \omega_o t \sin \omega_s t \end{aligned} \quad (6.4)$$

$$i = g_m \hat{E}_s \sin \omega_s t + \frac{k}{2} \hat{E}_o \hat{E}_s \cos (\omega_o - \omega_s) t - \frac{k}{2} \hat{E}_o \hat{E}_s \cos (\omega_o + \omega_s) t \quad (6.5)$$

A component having the frequency $f_o - f_s$ is present. This brings out the fact that any device that produces a product term such as the second term on the right-hand side of (6.4) will serve as a frequency converter.¹

The amplitude of the difference-frequency component in (6.5) depends directly upon \hat{E}_o ; therefore, under the conditions assumed, \hat{E}_o should be as large as possible in order to obtain a large converter output. However, the derivation no longer applies rigidly if the oscillator voltage is so large that g_m does not vary linearly with e_o . If the circuits under discussion were being used as amplitude modulators, the g_m variation would have to remain linear with respect to the modulating voltage (the modulating voltage replacing the oscillator voltage) in order to effect distortion-free modulation. However, in frequency conversion, nonlinearity between g_m and e_o is permissible and in fact is utilized to make the amplitude of the difference-frequency component appreciably larger.² Frequently the mixer is operated in a manner similar to Class B or C operation of an amplifier. There exists an optimum value for \hat{E}_o . When a remote-cutoff r-f pentode is employed in a single-input mixer circuit, the largest practical conversion transconductance (defined below) is obtained when the grid is biased as in Fig. 6.1c and is swung by the oscillator voltage e_o to the extent indicated. The signal-grid transconductance of a pentagrid mixer, Fig. 7.1c, is shown as a function of oscillator-grid potential in Fig. 6.1d. An oscillator-grid swing that results in large conversion transconductance and the corresponding values of g_m are shown. Converters frequently are operated with an oscillator voltage lower than that which produces maximum conversion transconductance, especially if the lower value improves the signal-to-noise ratio and reduces interference due to whistles.

Conversion transconductance is a mixer-tube coefficient. It is the ratio of the magnitude of the i-f carrier component of the plate

¹ H. STOCKMAN, A Treatment of Nonlinear Devices Based upon the Theory of Related Linear Functions, *J. Applied Phys.*, **16**, 645, December, 1943.

² E. W. HEROLD, The Operation of Frequency Converters and Mixers for Superheterodyne Reception, *Proc. I.R.E.*, **30**, 84, February, 1942. This article also contains a useful bibliography.

current to the magnitude of the r-f carrier voltage applied to the signal grid, with no load in the plate circuit and with constant plate voltage and is

$$g_c = \frac{|\hat{I}_{p(i-f)}|}{|\hat{E}_{g(r-f)}|} \Bigg]_{\substack{\text{for small amplitudes} \\ \text{with } e_0 \text{ const.}}} \quad (6.6)$$

In conventional converter circuits and at medium frequencies, g_c is of the order of a quarter to a third of the transconductance of the tube when used as an amplifier.

Conversion gain, or conversion amplification, is the ratio of the magnitude of the i-f carrier voltage developed across the converter load to the magnitude of the applied r-f carrier voltage. Strictly, the conversion gain is the limit of this ratio as the r-f voltage becomes smaller and smaller and may be written as

$$A_c = \frac{|E_{p(i-f)}|}{|E_{g(r-f)}|} \Bigg]_{\text{for small amplitudes}} \quad (6.7)$$

The voltage amplification of a circuit employing a tube having a large plate resistance, such as a pentode is practically equal to the product of the transconductance and the plate-load impedance, Chap. XIII, Sec. 12. Similarly, the conversion gain, or conversion voltage amplification, of a converter stage employing a tube having a high plate resistance is approximately

$$A_c \doteq |g_c Z_{i-f}| \quad (6.8)$$

where Z_{i-f} is the plate-load impedance at the intermediate frequency.

The signal and oscillator circuits may *interact* undesirably because of coupling between them. The coupling may affect the tuning of either circuit, Chap. XIV, Sec. 20, and may cause radiation at the oscillator frequency, especially if a preamplifier is not employed. Also, when interaction is present, an extremely strong incoming signal or a large disturbance voltage may alter the frequency of the oscillator and may even pull the oscillator into synchronism with the incoming signal. Interaction arising from circuit coupling is certain to occur in a single-input mixer but may also occur in a double-input circuit, perhaps through insufficient shielding. In a double-input circuit, interaction may result from *space-charge coupling*. The charge on the signal grid is affected by the surrounding space charge; and when the density of the space charge is forced to vary at the local-oscillator rate, alternating currents having the oscillator frequency are set up in the signal circuit. The effects of

space-charge coupling may be compensated in some cases by a lumped-circuit admittance connected between the oscillator and signal grids, placed inside or outside the envelope of the tube.

The frequency of the local oscillator should be as free as possible from drifts due to temperature variations. The oscillator should function with reasonable uniformity over the entire tuning range. The output voltage of the oscillator should be practically independent of variations in supply voltages and in the load. In general, the amplitude of the local-oscillator voltage employed in broadcast receivers is of the order of 10 or 20 volts.

In summary, the desirable characteristics of a superheterodyne frequency-converter stage are negligible interaction; large, uniform, and stable oscillator output; high signal-grid input impedance; high average mixer plate resistance; large and uniform conversion gain; large signal-to-noise ratio; good AVC action; and good "tracking" (constancy of difference frequency as the tuning dial is turned). In addition, low initial cost and low power consumption are desirable.

7. Superheterodyne Converter Circuits.—A variety of circuits may be employed for frequency conversion. An early circuit employing a triode and three circuits that employ multigrid tubes are discussed below. Converters for use at short and ultrashort wavelengths, employing diodes and crystals, are discussed in Sec. 14.

Early Triode Converter Circuit.—Figure 7.1a shows an early single-input mixer that is of interest because mixers that employ circuit coupling are now used in modern centimeter-wave receivers. The principal disadvantage of this type of converter is interaction. Also, the low plate resistance of the triode loads ("damps") the first i-f transformer, lowering the Q of the primary circuit, Chap. VII, Sec. 20, and reducing the selectivity. In addition, feedback through the grid-plate capacitance affects its operation adversely.

This type of converter circuit, when properly adjusted, has good sensitivity and conversion gain and has a very good signal-to-noise ratio especially when fixed-bias and plate-circuit rectification are employed. In this last respect the circuit is preferable to those employing multigrid tubes.

Pentagrid (Heptode) Converter.—Local oscillation and mixing are both accomplished in a *pentagrid-converter* tube by the use of a single electron stream. A circuit employing a pentagrid converter is shown in Fig. 7.1b. The cathode and grids 1 and 2 constitute a triode oscillator, grid 2 serving as a plate. The rest of the tube

is a tetrode. The oscillator grid (No. 1) modulates the electron stream, although the modulation is somewhat reduced by the opposite variation in potential of the oscillator plate (grid 2). The electrons in this passage from cathode to plate are acted upon by the oscillator voltage *before* the signal voltage, this being known as *inner-grid injection*. Grids 3 and 5 tied together serve as screen grids. To avoid grid current, the signal grid (No. 4) is operated with fixed negative bias, usually supplemented by AVC bias. When the potential of the oscillator grid is high, space charge builds up

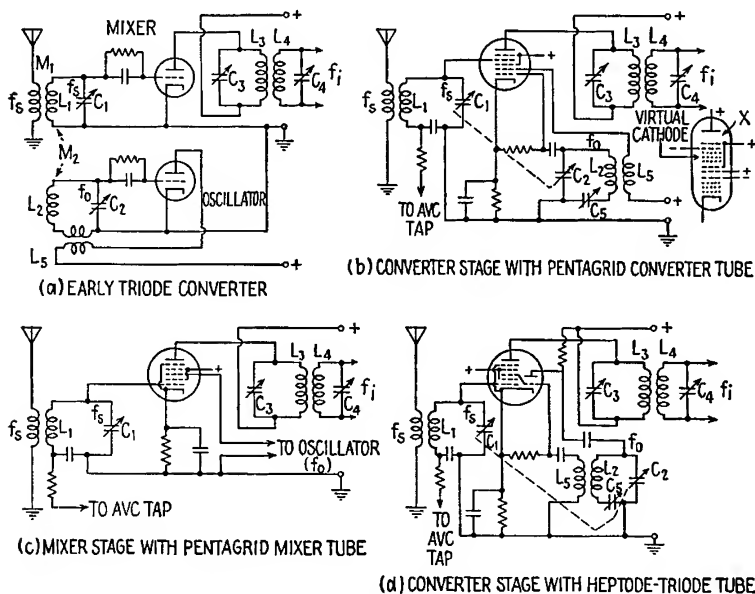


FIG. 7.1.—Superheterodyne mixer and converter circuits.

between grids 3 and 4, as shown at the right in the figure. The space charge becomes less dense as the oscillator-grid potential is reduced and may even become zero when the oscillator grid is sufficiently negative. The space charge serves as a virtual cathode for the tetrode portion of the tube. The tube becomes an *octode* if a sixth grid is added at the point marked *x*. This results in additional screening, increased average plate resistance, and less loading of the following i-f transformer.

Appreciable interaction may result from space-charge coupling. In addition, the "built-in" oscillator may operate improperly. These effects are most severe at ultra-high frequencies. Varia-

tions in space charge then cause detuning and other conditions unfavorable for proper operation. If the space-charge coupling is not properly neutralized, a large oscillator voltage may cause the signal grid to draw currents causing an appreciable direct current through the AVC filter resistances, upsetting the action of the AVC system.

The sensitivity and the conversion gain at medium frequencies are of the same order as in the triode circuit but are smaller at ultra-high frequencies. The signal-to-noise ratio is poorer than in the triode circuit because of the presence of so many grids. The signal-grid input impedance generally is much higher than in the triode circuit and may even be negative, raising the Q of the input circuit.¹ The heptode and octode load the i-f input transformer much less than a triode. These tubes are used widely in broadcast-band receivers because of the economy represented by combining the mixing and oscillator actions in one envelope.

Pentagrid (Heptode) Mixer.—The use of a separate local oscillator is preferable at higher frequencies. A circuit employing a *pentagrid-mixer* tube is shown in Fig. 7.1c. The signal is applied to grid 1. Grids 2 and 4, tied together, serve as screen grids, screening the oscillator grid (No. 3). Grid 5 is a suppressor grid, connected to the cathode. The oscillator voltage acts upon the electron stream *after* the signal grid, this being known as *outer-grid injection*. The signal grid has a remote-cutoff characteristic and usually is operated with AVC bias, while the oscillator grid has a sharp-cutoff characteristic.

Outer-grid injection results in less space-charge coupling than inner-grid injection, and therefore interaction is small. The conversion gain is high even in the short-wave region. Automatic-volume-control operation is possible down to a wavelength of several meters, below which fixed bias is preferred because AVC voltage then detunes the signal circuit. The pentagrid mixer, like the pentagrid converter, generally has a poorer signal-to-noise ratio than triode, diode, and crystal mixers.

Triode-hexode and Triode-heptode Converters.—As shown in Fig. 7.1d, a triode may be included in the same envelope with a hexode or heptode, the grid of the triode being connected internally to the hexode or heptode oscillator grid. Either inner-grid or outer-grid injection may be employed. The triode-heptode, which employs outer-grid injection, is practically free from interaction

¹ *Ibid.*

and can be used with good results and practically no radiation down to a wavelength of a few meters. Certain minor defects of this type of converter tube have been removed to some extent in recent designs by the use of close electrode spacing, guide electrodes, and internal shielding.

Oscillator Circuits.—A separate oscillator generally fulfills the requirements of an ideal oscillator better than the oscillator section of a triode-hexode or triode-heptode converter and considerably better than the pentagrid-converter oscillator. The built-in oscillator of the pentagrid converter sometimes exhibits a sort of motor-boating known as “flutter” which occurs because variations in supply and signal voltages may take place with aiding signs, making positive feedback possible. Frequency stability is much more important in ultra-short-wave converters than in medium-wave converters, since a deviation of 1 per cent in the oscillator frequency is much more serious in the former than in the latter. For this and other reasons a separate oscillator usually is employed at frequencies above approximately 50 mcps.

The frequency of the local oscillator may vary owing to change in temperature of its circuit elements. Temperature drift may be reduced by the use of temperature-compensated capacitors and similar devices. Variation in frequency also may be caused by changes in supply voltages, in the load on the oscillator, and in humidity (the latter affecting the values of the circuit elements).

Even a separate oscillator generally oscillates more strongly at one end of the tuning range. In the broadcast range, it may fail to oscillate at the low-frequency end, while parasitic oscillations may occur at the high-frequency end, producing whistles or squeals in the receiver output. At ultra-high frequencies it is more difficult to obtain approximately uniform operation over a broad tuning range. The operation of a conventional triode oscillator may be improved by altering the number of turns on the feedback coil, by changing the feedback coupling, and by adjusting the grid-leak and grid-capacitor values. Sometimes the uniformity of oscillator operation can be improved by placing resistors in the grid circuit close to the grid and in the feedback circuit. The oscillator should be designed so that it tracks properly.

8. The Intermediate-frequency Amplifier.—The i-f amplifier is a high-gain high-frequency amplifier designed to operate at a fixed frequency. Its function is to amplify the weak i-f output of a converter to a voltage sufficient to operate the detector, without

introducing undesirable frequency, phase, or nonlinear distortion. Circuit diagrams that include single-stage i-f amplifiers are shown in Figs. 5.2 and 11.1.

As stated in Sec. 5, the shape of the frequency-response curve and the width of the pass band of the receiver as a whole may be controlled readily by controlling the band-pass characteristics of the i-f amplifier. The i-f transformers may be designed to have a more or less well-defined double peak in their frequency-response characteristics, Chap. VII, Sec. 15. This reduces the possibility of side-band clipping, improves the reproduction of the higher audio or video frequencies without loss of adjacent-channel selectivity, and reduces the stability required of the local oscillator. On the other hand, the use of double-peaked transformers increases the cost of manufacture and of alignment and testing. Furthermore, the anticipated improvement in fidelity is not realized unless the detector, the low-frequency amplifier, and the output device also have high-fidelity characteristics. The wider the pass band, the more free the other parts of the receiver must be from nonlinear and phase distortion. For these reasons, transformers having a single-peak frequency-response curve are used in small and inexpensive receivers. Wide-band reception is not always desirable because the background noise in the output increases with the band width.

9. Image Interference and Its Reduction.—The frequency response of a superheterodyne whose r-f input circuit is tuned to a

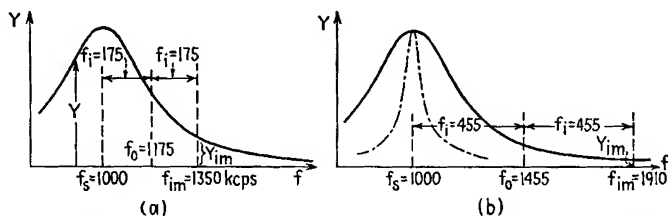


FIG. 9.1.—Image-frequency response in a converter: (a) intermediate frequency = 175 kcps; (b) intermediate frequency = 455 kcps.

frequency f_s of 1,000 kcps is indicated in Figs. 9.1a, b. First, assume that the oscillator frequency f_o is 1,175 kcps and that the intermediate frequency f_i is 175 kcps (a value used in earlier designs). If there should be an incoming transmission at the frequency $f_{im} = 1,350$ kcps, Fig. 9.1a, it will cause a difference frequency of $1,350 - 1,175$ or 175 kcps in the converter stage and will be heard in the background, perhaps only as a whistle. Such an undesired station is

known as an *image station* or *image*, and the frequency $f_{im} = f_s + 2f_i$ is known as the *image frequency*.¹ The *image ratio* of a receiver is the ratio of the input required at the frequency f_{im} to that required at the frequency f_s to produce output signals of the same strength. The image ratio must be high for image-channel interference, or image interference, to be negligible.

To secure a high image ratio the intermediate frequency is made as large as other considerations permit. An intermediate frequency of 455 kcps is used frequently in modern broadcast receivers. As seen in Fig. 9.1b, the admittance of the receiver at the image frequency of 1,910 kcps is extremely small. However, an increase in intermediate frequency generally reduces the selectivity and the stability of the receiver. Assuming coils of the same Q , the relation $Q = f_i' / (f_i'' - f_i')$ indicates that the half-power band width $f_i'' - f_i'$ of the i-f amplifier is several times greater for an intermediate frequency of 455 kcps than for 175 kcps.

Image interference also is reduced by increasing the r-f input selectivity of the receiver. If the r-f input response curve is made sharper, as indicated by the dot-dash curve in Fig. 9.1b, the image ratio is increased. However, the response curve of the input circuits must not be made too narrow, for side-band clipping may then be encountered. The selectivity is obtained from the circuits, not the tubes, so that any practical amount of preselectivity can be obtained in general without the use of additional tubes. The tube or tubes in a preamplifier in themselves do not contribute to the image-suppression qualities of a preamplifier.

10. Preamplifier, Detector, Low-frequency Amplifier.—The *preamplifier* in a conventional superheterodyne is a one- or two-stage amplifier with variable tuning, in which AVC-controlled remote-cutoff pentodes usually are employed. The preamplifier may be included in the receiver for one or more of the following purposes:

1. To eliminate image-channel and similar types of interference.
2. To minimize adjacent-channel interference.
3. To increase the signal-to-noise ratio of the receiver.
4. To prevent radiation by isolating the local oscillator from the antenna.
5. To make the receiver less dependent upon the antenna characteristics.
6. To increase the over-all gain of the receiver.

¹ If f_o is less than f_s , the image frequency is given by $f_{im} = f_s - 2f_i$.

Ordinarily, image-channel suppression is a major function of the preamplifier. However, at higher frequencies image suppression may not be required because there may be no radio signals at the image frequency. Adjacent-channel selectivity is not usually the primary function of the preamplifier because the i-f amplifier is far more effective in this respect. Prevention of radiation may be of great importance in special designs for military use but is of less importance in commercial designs, where radiation is usually insufficient to cause noticeable interaction between neighboring receivers. Variation in the impedance of the antenna over the tuning range is seldom of such importance that it alone justifies the introduction of a preamplifier. At present, the contribution of the preamplifier tube to the over-all gain of a receiver is small in the ultra-short-wave band, so that at the higher radio frequencies the preamplifier is not as effective as at the lower frequencies.

The *detector* in a superheterodyne receiver may be of almost any type. Except in small "personal" radios and similar receivers, a large-signal detector commonly is employed. The detector circuit most commonly used is the half-wave diode circuit, since it distorts the signal negligibly except at high percentages of modulation and can also supply control voltage for an AVC system, as in Fig. 5.2.

The *low-frequency amplifier* raises the level of the output of the detector sufficiently to operate a loudspeaker or a picture tube. A manually operated volume or gain control, usually located at the input to the low-frequency amplifier, is required even in a set that includes a conventional AVC system. Even though the AVC system may hold constant the amplitude of the carrier that reaches the detector, differences in the degree of modulation of the various signals applied to the detector cause comparatively large variations in the magnitude of its low-frequency output.

11. Automatic Volume Control.—Nearly all modern superheterodyne receivers include either an automatic-volume-control (AVC) system or a delayed automatic-volume-control (DAVC) system, to keep the output of the receiver at a fairly constant level, preventing "blasting" of the loudspeaker or the picture tube. The conventional AVC systems maintain a more or less constant i-f carrier input to the detector by decreasing the gain of the high-frequency amplifier as the amplitude of the incoming carrier increases.

A commonly used AVC system is shown in Fig. 5.2. Voltage

developed by rectification across the detector load is supplied to an AVC line through a filter $R_{10}C_{20}$, which is intended to attenuate the i-f and a-f voltages sufficiently so that they have negligible effect on the grid bias. The AVC line supplies the r-f amplifier, converter, and i-f amplifier with AVC bias voltage through individual decoupling filters, such as R_1C_2 . The AVC bias voltage is additional to a more or less fixed bias such as cathode bias. This more or less fixed bias determines a suitable quiescent point to prevent grid current and excessive plate current in the absence of a signal. The greater the carrier amplitude of the i-f wave that reaches the detector, the greater the AVC bias voltage and the smaller the gain of the AVC-controlled stages.

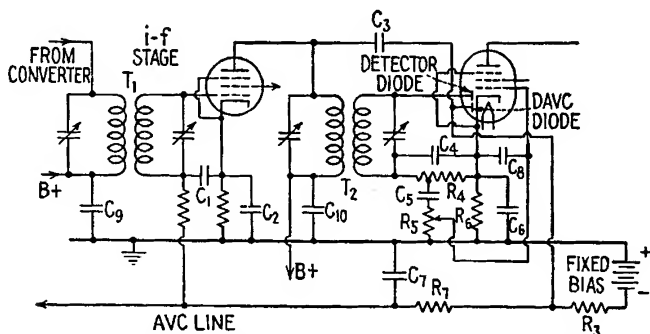


FIG. 11.1.—Conventional i-f and duplex-diode-pentode stages in a superheterodyne receiver with delayed automatic volume control.

Improved AVC action is obtained by using a separate rectifier coupled by a capacitor to the plate of an i-f amplifier, Fig. 11.1, or to the plate of the detector. Intermediate-frequency and d-c amplifiers also may be incorporated in the AVC system for improvement in its operation.

The choice of values of resistance and capacitance employed in an AVC filter system is a matter of reaching a difficult compromise. The lowest modulation tones are difficult to filter out, and considerable negative feedback may occur at these frequencies in small and inexpensive receivers, reducing the already poor base response. A large time constant causes the lowest modulation tones to be filtered out but slows down the AVC action. If a large time constant is obtained by use of large values of resistance in the AVC line, a very small amount of grid current causes an appreciable drop across the resistances, which usually makes the controlled grids more negative and results in poor AVC regulation. Likewise, the

use of large capacitors is not consistent with good design. Variations in the input signal due to fading and resetting of the tuning dial are compensated reasonably well in a broadcast receiver having an AVC time constant of the order of 0.1 sec.

Ordinary AVC reduces the sensitivity of a receiver even when the incoming signal is weak, which is a disadvantage. In a DAVC system, AVC bias voltage is not developed until the strength of the incoming signal exceeds a predetermined value. An example of a DAVC system is shown in Fig. 11.1. C_3 serves as coupling capacitance and R_3 as DAVC diode-load element, while R_7 and C_7 constitute a filter in the AVC line. The cathode of the duplex-diode-pentode is maintained a few volts above ground potential by means of cathode bias. A fixed bias holds the AVC line a few volts below ground potential. Therefore, in the absence of an i-f signal, the plate voltage of the DAVC diode is the sum of the cathode bias across R_6 and the fixed bias, and the DAVC diode is not conducting.

When the potential of the DAVC plate is swung at the i-f rate with an amplitude that is small and is not sufficient to bring the plate up to the point of conduction, there is no change in the AVC line potential. When the amplitude is sufficient to cause conduction of the diode during the positive peaks, pulses of current occur in R_3 and the potential of the AVC line with respect to ground is made more negative. The pulsating voltage drop across R_3 is smoothed by the filter R_7C_7 . The critical amplitude (the amplitude at which conduction begins) is slightly less than the sum of the cathode-bias and fixed-bias voltages.

When a receiver is equipped with either AVC or DAVC, ordinary fading causes a rise and fall in the level of the background noise, which means that noise replaces the signal when the signal fades. Selective fading, which is unequal fading of the various frequency components in the incoming modulated wave, is made especially evident. If one side band disappears and the carrier is weakened, the sensitivity of the receiver increases and the remaining part of the transmission produces a large and distorted output signal. The sensitivity of a receiver that has AVC rises when the receiver is tuned between stations. The background noise may become very prominent between stations unless another control circuit is employed to suppress this noise.

By extending the AVC regulation to the low-frequency amplifier, the output of a receiver can be held at a fixed level independent of the degree of modulation of the incoming signal. Generally

this is not considered worth the cost in broadcast receivers. Over-regulation, which may be undesirable, becomes possible when AVC is extended to the low-frequency amplifier or to any stage following that from which the AVC system derives its r-f controlling signal.

12. Automatic Frequency Control.—Inaccurate tuning, frequency drift in the local oscillator, and error in alignment may cause the frequency applied to the i-f amplifier to be incorrect. An automatic-frequency-control (AFC) system is incorporated in some receivers to correct the frequency of the i-f signal.¹ Such a system is especially valuable in push-button-tuned or remotely controlled receivers. Automatic frequency control facilitates the

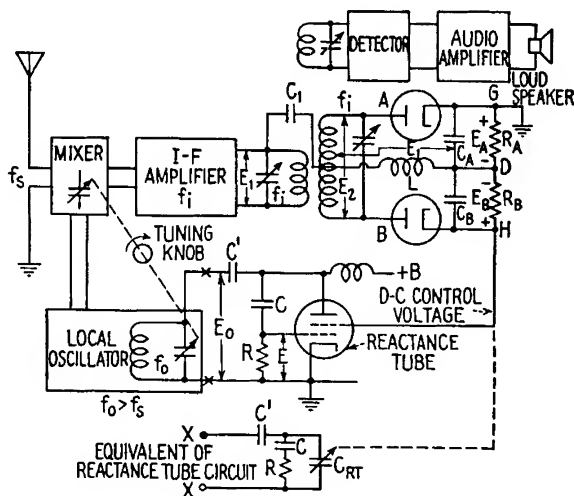


FIG. 12.1.—Automatic-frequency-control (AFC) system.

tuning by reducing the accuracy with which the tuning dial must be adjusted.

An example of an AFC system is shown in Fig. 12.1. A reactance-tube circuit, Chap. XX, Sec. 10, controls the frequency of the local oscillator. The reactance tube (in this example) simulates a variable capacitance C_{RT} , connected across the oscillator tank circuit. A direct control voltage is applied to the reactance tube from a discriminator, Chap. XXI, Sec. 22. The primary and secondary of the discriminator transformer are tuned to the correct intermediate frequency. The output of the i-f amplifier is applied to both the detector and the discriminator.

¹ J. F. RIDER, "Automatic Frequency Control Systems," J. F. Rider, Publisher, 1937.

When the carrier frequency of the i-f wave has the correct value, the output voltage of the discriminator is zero. If the carrier frequency of the i-f wave should be above the correct value, the discriminator develops and applies to the reactance tube a d-c voltage that in this example is positive. The positive voltage increases the capacitance simulated by the reactance tube, thus lowering the frequency of the oscillator, and in turn lowering the carrier frequency of the i-f wave until the error in frequency practically is eliminated. When the carrier frequency of the i-f wave is below the proper value, the system operates in the opposite direction. Frequency control is possible as long as the carrier frequency of the i-f wave remains within the more or less linear range of response of the discriminator. An AFC system generally can correct an error of several kilocycles in a 455-kcps i-f amplifier.

SHORT-WAVE RECEIVERS

13. General.—Unconventional features of construction are necessary in short-wave receivers. At frequencies of the order of hundreds of megacycles, tuned sections of transmission line frequently are used as resonant circuits, while cavity resonators are employed at still higher frequencies. The requirements regarding wiring, switching, and shielding are severe. Tubes and tuning elements must be wired together with considerable care in order to avoid undesired admittances. Standard types of short-wave receivers may employ the same types of tubes found in medium-wave receivers. Frequently, special types of tubes, for example “acorn” tubes, are used. Special tubes for amplification, conversion, detection, and oscillation have been developed for use in the more unconventional centimeter-wave receivers.

Regenerative-detector circuits operate with reduced sensitivity and selectivity at high frequencies, while superregenerative detectors, up to a certain point, improve in sensitivity with frequency. For this and other reasons the superregenerative detector frequently is preferred to the regenerative detector at ultra-high frequencies. However, for selectivity or for some other reason, a superheterodyne circuit may be preferred.

Short-wave and ultra-short-wave stations generally are grouped in frequency bands. Each tuning range in a short-wave receiver may cover two or more of these bands, but in ultra-short-wave receivers a separate tuning range may be provided for each band. For tuning within a given frequency band, a system known as

“band spreading” (mechanical, electrical, or both) has been developed. This is adequately described in the literature. The design of tuning systems in centimeter-wave receivers is a complicated problem for which ordinary band spreading is not an adequate solution.¹

14. Short-wave Superheterodyne.—The short-wave superheterodyne receiver differs from the medium- and the long-wave superheterodyne mainly in the design of the frequency converter and its associated circuits.² In some cases two converters are employed, the carrier frequency being changed twice before detection. A preamplifier generally is not employed because at ultra-high frequencies a preamplifier may have a power gain less than

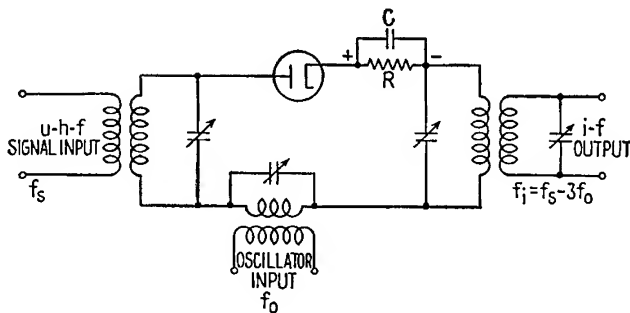


FIG. 14.1.—Diode-mixer circuit for u-h-f operation. Coil-capacitor arrangements on input side shown merely as symbols.

unity, may produce more noise power than signal power, and may not be required for image suppression.

Diodes and crystals frequently are used as mixers at frequencies of the order of thousands of megacycles, because the types described in Sec. 7 do not perform adequately at these frequencies. Electron coupling is not used because of transit-time effects and other limitations. At progressively higher frequencies, transit-time effects render first the triode, then the diode inefficient, and the sphere of usefulness of the crystal then is entered. Factors other than efficiency, however, are included in the choice of mixer. Diode and crystal mixers have all the limitations of circuit-coupled mixers

¹ S. Y. WHITE, The Precision Tuning Problem in U-h-f Broadcasting, *Electronics*, **16**, 94, May, 1943.

² E. W. HEROLD and L. MALTER, Some Aspects of Radio Reception at Ultra High Frequency, *Proc. I.R.E.*, **31**, 423; **31**, 491; **31**, 567; August, September, October, 1943.

discussed in Secs. 6 and 7, including the disadvantage of strong radiation.

Diode Mixer.—A simple u-h-f diode-mixer circuit is shown in Fig. 14.1. Coils and capacitors are used merely as symbols on the input side of the circuit, since tuned lines and cavity resonators frequently are employed. The local-oscillator voltage preferably should be of the order of several volts. Rectification of the local-oscillator voltage produces a bias voltage across the combination RC that may amount to several volts. Therefore, the diode conducts only during the positive peaks of the oscillator voltage, and its operation resembles Class C operation of an amplifier tube. Contrary to the practice customary in respect to broadcast receivers, the local oscillator here usually is operated at a frequency

$$f_o = f_s - f_i,$$

that is, below that of the incoming signal, in order to obtain greater stability. Considerable improvement in operation may be obtained in some cases by employing a multiple of the oscillator frequency in the determination of the difference frequency; for example, by letting $f_i = f_s - 3f_o$. Then when the circuit for injection of the oscillator voltage is tuned to the frequency f_o , the impedance presented by this circuit to the signal current is much reduced, which is an advantage. However, the conversion power ratio of the mixer is also reduced, which is a disadvantage.

Note that, when an integral multiple of the oscillator frequency is employed in frequency conversion, the waveform of the local oscillator may be purely sinusoidal;¹ for nonlinearity in the *transconductance* is the necessary requirement, and this is contributed by the mixing device, which is a nonlinear load on the oscillator.

Special attention must be given to impedance matching in the design of a u-h-f converter because the conversion power ratio is of greater importance than the conversion gain of the mixer.² The conversion gain of nearly all u-h-f converters is less than unity. To some extent the attenuation of the signal may be compensated by additional gain in the i-f amplifier.

Crystal Mixer.—A rectifier crystal, for example a silicon-tungsten cartridge, may be used in place of the diode in a mixer circuit of the

¹ HEROLD, *loc. cit.*

² E. G. JAMES and J. E. HOULDIN, Diode Frequency Changers, *Wireless Engineer*, **20**, 15, January, 1943; F. M. COLEBROOK and G. H. ASTON, Diode as a Frequency Changer, *Wireless Engineer*, **20**, 5, January, 1943.

type shown in Fig. 14.1. A better signal-to-noise ratio thus may be obtained. The conversion gain, compared with that of a diode, may be either greater or less, depending upon a number of factors. The average resistance of crystals is smaller than that of diodes, so that the i-f transformer must have a larger step-up ratio. Because crystals conduct to some extent in the "reverse" direction, Class C operation and the use of a multiple of the oscillator frequency may not be possible. In a crystal mixer, the maximum conversion power ratio may be obtained when the oscillator voltage is limited to the order of tenths of a volt.¹

Other U-h-f Mixers.—In theory, any device that operates non-linearly may be used for frequency conversion. Several special u-h-f oscillators have been developed, and there are possibilities for mixing by using the oscillator circuit, employing the so-called "autodyne" principle.² These processes are of interest because of the difficulties in providing external mixing circuits at ultra-high frequencies.

15. Frequency-modulation Receivers.—A frequency-modulation receiver generally is a superheterodyne. In many cases receivers are designed for both frequency-modulation and amplitude-modulation reception. A block diagram of a conventional receiver for both frequency-modulation and amplitude-modulation is shown in Fig. 15.1. Either one of the detector systems can be switched in between the i-f and a-f amplifiers. The frequency-modulation detector includes a discriminator. Frequency-modulation signals are passed through one or two limiter stages to remove any variation in amplitude that the modulated wave may have acquired in transmission and reception. The detector circuit usually is followed by a deemphasis circuit to attenuate the higher modulation tones, these tones usually having been emphasized at the transmitter. The first two i-f amplifier stages have a relatively broad frequency response. Therefore an added narrow-band i-f stage is employed ahead of the amplitude-modulation detector to obtain the greater selectivity required in amplitude-modulation reception.

The primary purpose of a preamplifier, when used in a frequency-modulation receiver, is usually to increase the signal-to-noise ratio.

¹ H. STOCKMAN, UHF Converter Analysis, *Electronics*, **18**, 140, February, 1945.

² J. F. RIDER, "Servicing Superheterodynes," 3d ed., pp. 90-99, J. F. Rider, Publisher, 1942.

It may also be used for image suppression, although image interference usually is not a major problem.

In a frequency-modulation receiver for operation in the 42 to 50 mcps broadcast band, the converter may be a pentagrid mixer with separate oscillator or a modern triode-hexode or triode-heptode. In a receiver for operation at frequencies of several hundred or more megacycles/second, tubes of the acorn type may be employed in the high-frequency part of the receiver.

To accommodate the side bands of the frequency-modulation wave, the i-f amplifier must have a broad frequency response. The required band width may range from 30 kcps for police or similar

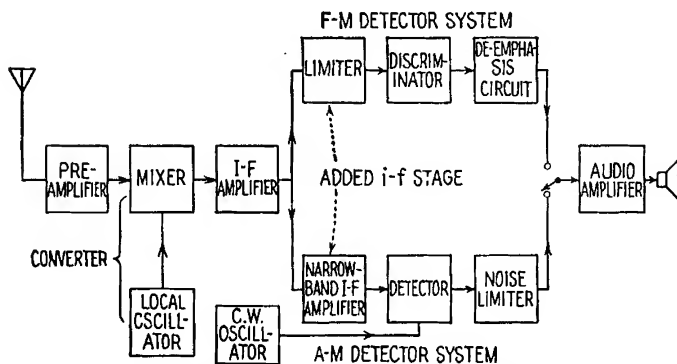


FIG. 15.1.—Block diagram of a conventional communication receiver for both amplitude and frequency modulation.

transmissions up to 200 kcps for broadcasting. To achieve the necessary band width, the coupling transformers may be designed to have band-pass characteristics and may be shunted with resistors. In some receivers, “staggering” of the circuits may be employed, the frequencies of the individual circuits being grouped around the carrier frequency. Two i-f stages generally are required in order to amplify the signal sufficiently to operate the limiter properly. The considerations that govern the choice of the intermediate frequency are essentially the same as in an amplitude-modulation receiver. In prewar frequency-modulation broadcast receivers, the intermediate frequency is of the order of 3 to 5 mcps. The intermediate frequency is still higher in ultra-short-wave frequency-modulation receivers.

Essentially, a frequency-modulation limiter stage is a considerably overdriven i-f amplifier stage. Its voltage amplification is of the order of unity, and the amplitude of its output is of the order of

10 volts. An r-f pentode with a sharp-cutoff characteristic generally is employed. Limiting action usually is achieved by the use of a resistor in the grid circuit and reduced plate and screen-grid potentials. The bias voltage developed in the grid circuit by large signals, together with the reduced polarizing potentials, results in clipping. Detailed analysis reveals the presence of a number of interrelated effects that contribute to the final result.¹ The time constant of the grid coupling capacitor and grid resistor should be small, of the order of a few microseconds, so that rapid variations in amplitude, such as those characteristic of impulse noise, may be eliminated.² A two-stage limiter offers the possibility of improved limiting action, partly because the two stages can be given different time constants.

The discriminator employed as detector is of the types discussed in Sec. 12 and in Chap. XXI. The time constant of the deemphasis circuit,^{2,3} is 100 microsec for the 42 to 50 mcps broadcast band and 75 microsec for the 88 to 108 mcps broadcast band. The quality of the a-f amplifier and of the loudspeaker usually is higher than in the average amplitude-modulation short-wave receiver, and may be designed to cover a frequency range from 50 to 15,000 cps. The audio output stages are usually of the push-pull type, and sometimes separate treble and bass loudspeakers are employed.

The excellent tone quality obtainable in frequency-modulation reception may be achieved in some cases with amplitude-modulation reception. The primary advantage of frequency-modulation reception lies in the low noise level obtained even under adverse conditions of reception. The low noise level permits an extended range of loudness.

DISTURBANCE TO RADIO RECEPTION

Disturbances to radio reception occur as hissing, crackling or crashing sounds, hum or whistles, undesired transmissions; to video reception, as an imperfect image. The signal power at the reproducer must exceed appreciably the disturbance power in order to secure satisfactory reproduction of the original intelligence. Import-

¹ J. A. WORCESTER, JR., Recent Improvements in Frequency Modulation Receiver Design, *RMA Tech. Bull.* 2, Nov. 12, 1940.

² A. HUND, "Frequency Modulation," pp. 249-264, 204-225, McGraw-Hill Book Company, Inc., 1942.

³ Chap. XX, Sec. 9.

tant types of radio disturbance are fluctuation noise, impulse noise, and radio-station interference.

16. Fluctuation Noise.—Random electrical variations of a character that cause a hiss without definite pitch in the loudspeaker are known as fluctuation-noise voltages. These voltages originate in tubes and other circuit elements and come into being because electric charge exists in distinct quantities instead of in a continuum. Generally, the frequency spectrum of fluctuation-noise voltages is continuous and has a height that in a given receiver depends upon the frequency-response characteristics of the circuits. The fluctuation-noise voltages of most interest are those due to shot effect and to thermal agitation.

Shot Effect.—The time average of the number of electrons reaching the plate is subject to random variations because emission of electrons from the cathode is not absolutely steady. The irregularity in the plate current due to this cause is known as shot effect. When a tube is operated at a plate potential such that space charge is developed, the space charge serves as a reservoir or cushion, smoothing out the flow of electrons and reducing the shot effect.

Thermal Agitation.—Thermal-agitation voltages¹ are small fluctuating differences of potential that appear between the terminals of a conductor or circuit owing to random motion of electrons within it, the motion of the electrons being a concomitant of the heat energy of the material. The effect of the thermal agitation is equivalent to the effect of an emf (in series with the conductor considered noise-free) given by

$$E^2 = 4kT \int_{f_1}^{f_2} (R)_f df \quad (16.1)$$

where E is the rms value of the fluctuating emf in volts, k is Boltzmann's constant ($1.37 \cdot 10^{-23}$ joules/°K), T is the absolute temperature of the conductor, f_1 and f_2 are the limits of the range of frequency that is of interest, and $(R)_f$ is the resistive part of the impedance of the conductor at the frequency f .

The value of $(R)_f$ for a parallel tuned circuit, when plotted against frequency, forms a peaked curve. The thermal-agitation voltage contributed by a tuned circuit may be determined by integrating (16.1) over the relatively narrow band of frequencies covered by the circuit.

¹ J. B. JOHNSON, Thermal Agitation of Electricity in Conductors, *Phys. Rev.*, **32**, 97–109, July, 1928; H. NYQUIST, Thermal Agitation of Electric Charge in Conductors, *Phys. Rev.*, **32**, 110–113, July, 1928.

When $(R)_f$ is independent of frequency, (16.1) reduces to the simpler form

$$\begin{aligned} E &= 2 \sqrt{kTR(f_2 - f_1)} \\ &= \sqrt{5.5 \cdot 10^{-23} TR(f_2 - f_1)} \end{aligned} \quad (16.2)$$

This equation may be used, for example, to calculate the noise voltage contributed by the input grid resistor of an amplifier, the range of frequency employed being the range of frequency covered by the amplifier.

When the plate current of a tube is limited by space charge, the residual noise resulting from shot effect may be evaluated by use of (16.2) as a thermal agitation noise in the plate resistance r_p of the tube. The "temperature" of the plate resistance is chosen to yield the correct value of the noise.

Other Sources.—Additional sources of fluctuation noise within a tube are ionization, leakage currents, minute grid currents, and in multielectrode tubes the random division of current among the various electrodes. This random division contributes largely to the noise inherent in multielectrode tubes.

Fluctuations occur in every circuit and circuit element in both transmitter and receiver, but the level of the signal in the transmitter is usually far above the level of the disturbances. Fluctuation voltages usually are troublesome only in the first stage and perhaps in the second stage of a sensitive receiver or amplifier. In these stages the signal voltage may be of the same order of magnitude as the fluctuation voltage. To reduce fluctuation noise due to shot effect, the first tube and perhaps the second tube may be operated with reduced plate voltage.

17. Impulse Noise.—Impulse noises are disturbances of steep-wavefront character and may be either natural or man-made. Impulse noises may be picked up by the antenna system in addition to the desired signal, may enter the receiver directly, or may enter it by way of the power line.

Impulse disturbances have very broad energy spectra of the types shown in Chap. XIX, Figs. 13.1*e, f, g*. The spectrum of the desired signal sets a minimum limit upon the width of the response curve, and therefore a certain amount of impulse disturbance must be permitted to enter the receiver.

Atmospheric disturbances ("static") occasionally are so intense that, at long and medium wavelengths, reception is impossible. Static usually is less intense at medium and short wavelengths and

is practically nonexistent above 50 to 100 mcps. Atmospheric disturbances sometimes resemble fluctuation noise instead of impulse noise. Numerous devices intended to "eliminate" static, *i.e.*, prevent it from causing any disturbance in the receiver, have been patented. Experience has indicated that static of a random nature cannot be eliminated. Improved reception has been obtained, however, by means of increased antenna power at the transmitter, directional antennas, diversity reception systems, etc. A satisfactory solution in some cases is the use of angle modulation (frequency modulation).

Man-made impulse disturbances originate principally from electrical appliances, high-voltage power lines, electric railways, automobile and aircraft electrical ignition systems, and the like. In many cases these disturbances can be suppressed or eliminated at the source by means of filtering and shielding. Ignition disturbances may be very intense in the region of 50 to 100 mcps but, as with static, usually do not impair reception above 100 mcps. Therefore above 100 mcps it usually is noise originating within the receiver that limits its effectiveness.¹

Impulse disturbances that enter amplitude-modulation receivers, especially code receivers, sometimes are suppressed by means of a noise-limiter circuit, which may be an amplitude-limiting stage following the detector, or a silencer circuit that shuts off or alters the amplification of the i-f amplifier for the duration of each disturbance impulse.²

18. Interference.—Unwanted transmissions and whistles heard in addition to the desired transmission, and due to the existence of stations other than the desired one, are referred to as radio-station interference or merely as "interference." Radio-station interference may be due to insufficiently selective circuits in the receiver or to some improper action within the receiver itself. Interference due to the first cause may be remedied by increasing the Q of coils and capacitors, employing coupled circuits, and similar means. This improvement generally is most effective if carried out in the i-f amplifier, Sec. 8. By use of a quartz crystal in the i-f amplifier, a half-power band width as small as a few hundred cycles/second

¹ B. DUDLEY, U.H.F. Reception and Receivers, *Electronics*, **15**, 51, April, 1942.

² A.R.R.L. "Radio Amateurs' Handbook," any recent edition; consult index for noise reduction. C. WASMANSDORF, Reducing Radio Noise, *Electronic Ind.*, **3**, 80, July, 1944.

(suitable only for code reception) may be obtained. Where a quartz crystal is thus employed, it is also possible to "phase out," or reject completely, a particular interfering CW station.¹

The following types of radio-station interference result from improper action in a receiver:

1. Cross-talk interference. If a receiver is tuned to a station and if the first tube (and possibly the second) operates on a nonlinear portion of its characteristic, another strong station may produce an emf in the grid circuit so that sum and difference terms appear as side bands of the carrier of the desired transmission. Thus the modulation characteristics of the undesired station may reach the detector. This is known as cross-talk interference or cross-modulation and may be reduced by increased selectivity in the r-f input circuits and also by the use of "supercontrol" tubes whose i_b - e_c characteristics are not sharply bent at any point.

2. Harmonic conversion interference. An incoming transmission that has a carrier frequency $2f_o \pm f_i$, for example, may result in conversion based upon the second integral multiple of the local-oscillator frequency, creating an i-f signal having the modulation characteristics of this incoming transmission.

3. Intermediate-frequency separation interference. Two transmissions whose difference in frequency equals the intermediate frequency may produce an i-f signal.

4. Intermediate frequency subharmonic conversion interference. When an incoming transmission produces a strong difference frequency of value $f_i/2$ or $f_i/3$, etc., the second or third harmonic components of these difference frequencies will be at the intermediate frequency and may create an i-f signal.

5. Image interference. This important type of interference has been discussed in Sec. 9. The four types of interference 2 to 5, which originate in the converter, may be remedied by increased selectivity ahead of the converter.

6. Detector harmonic interference. In some receivers the detector or the i-f amplifier may produce components of frequency $2f_i$, $3f_i$, $4f_i$, etc., which, though very weak, may be strong in comparison with the incoming wave and may be picked up by the input circuits. This may be remedied by shielding the detector circuit and filtering its output, thus reducing the coupling between the detector

¹ A.R.R.L. "Radio Amateurs' Handbook," any recent edition; consult index for crystal filters.

and the earlier circuits, or by employing a detector that does not produce appreciable harmonics.

A whistle may be caused by the carrier wave of an interfering station through one of the processes just listed, even though the modulation of the interfering station is too weak to be perceived. Whistles may also be sum or difference tones resulting from spurious or parasitic oscillations in the receiver.

19. Other Disturbances.—*Residual hum* is the hum that remains when the low-frequency amplifier and power supply are in operation with the high-frequency part of the receiver disconnected from the detector. Residual hum may be caused by insufficient filtering in the power supply, induction of hum-frequency voltages in the amplifier, or leakage between heater and cathode in tubes. *Modulation hum* results from modulation of the carrier in the r-f or i-f stages at the hum frequency.

Mechanical vibration of tubes, coils, capacitors, etc., may result in currents having the vibration frequency or currents modulated at the vibration frequency. This effect is known as *microphonic action*, and the noise as *microphonic noise*.

20. Radiation from Receivers.—As stated in previous sections, normal or abnormal oscillations within a receiver may be radiated by the antenna system. To reduce the radiation, preamplifier stages or balanced circuits may be used to isolate the oscillations from the antenna system.

A receiver may also radiate directly or may radiate by way of the power line. Direct radiation may be reduced by shielding. Power-line radiation may be reduced by the use of suitable line filters and by shielded transformers.

RECEIVER MEASUREMENTS AND SERVICING

21. Measurement of Sensitivity, Selectivity, and Fidelity.—The characteristics of a medium-wave or a long-wave receiver usually are determined by the use of a standard-signal generator and an output measuring device. Various techniques of measurement may be employed; there is no standard method that is applicable to all receivers. The following techniques are based upon the I.R.E. Standards on Radio Receivers, 1938 and were formulated to apply to receivers of that date.

For measurement of sensitivity, a signal that is amplitude-modulated 30 per cent with a 400-cycle tone is applied to the receiver. The amplitude of the applied signal is adjusted until a

specified receiver output signal is obtained. The sensitivity usually is expressed in terms of the input signal strength, in microvolts, required to produce a specified output signal.

During measurement of the selectivity of a receiver not equipped with AVC, a test signal is applied at the frequency to which the receiver is tuned, and the amplitude of the applied signal is adjusted until a specified receiver output signal is obtained. Then, without readjusting the receiver in any way, the signal generator is detuned in steps on both sides of resonance, and at each step the signal voltage is readjusted until the specified receiver output signal is again obtained. The ratio

$$\left. \frac{\text{Off-resonance input voltage}}{\text{Input voltage at resonance}} \right]_{\text{at constant output}}$$

increases as the frequency of the applied signal departs from resonance. This ratio may be plotted (on a logarithmic scale) as a function of the frequency of the test signal, and the resultant curve is known as a selectivity curve.

If a receiver is equipped with AVC, its selectivity may be measured with the AVC system disconnected. For more rigorous measurement with the AVC system in operation, a so-called "two-signal" test may be made as described in the I.R.E. standards.

For measurement of fidelity, the signal generator is adjusted to the frequency to which the receiver is tuned, the audio frequency of modulation of the test signal is varied, the percentage of modulation being left constant, and the electrical or acoustical output of the receiver is measured. The results may be plotted in a manner similar to that for the frequency response of an amplifier.

The techniques described may be modified and extended to apply to television, frequency-modulation and other types of receiver.

22. Measurement of Signal-to-noise Ratio.—Measurement of the disturbance content of a receiver output signal, in order to determine the signal-to-noise ratio, is difficult. The deflections, due to noise, of an instrument attached across the receiver output terminals are in some cases difficult to interpret. In some more recent methods, the noise content is measured by the use of a specially designed noise meter connected at the output of the i-f amplifier.

In short-wave receivers it is possible and desirable to regard the antenna as an integral part of the receiving system, and the

total characteristics of the system then may be used to judge its usefulness.¹ The disturbance originating within the receiver proper may be evaluated as an "equivalent receiver-noise emf" at the antenna terminals, so that in effect there is connected between the antenna terminals the series combination of the signal-field emf, the disturbance-field emf, and the terminal impedance of the antenna. Then, the total signal-to-noise ratio of the receiving system may be defined as the ratio of the signal-field emf to the sum of the disturbance-field emf and the equivalent receiver-noise emf. Receiver characteristics thus may be compared on an absolute basis.

When noise is not considered, the principle of maximum power transfer is employed in matching the antenna and receiver impedances. When noise is considered, conditions for maximum power transfer do not always coincide with conditions for maximum signal-to-noise ratio and a compromise must be made. In a receiver that operates a cathode-ray tube, for example, the signal-to-noise ratio may be just as important as the selectivity and the fidelity.

23. Fundamentals of Servicing and Alignment.—The important step in servicing an inoperative receiver is locating the fault in a minimum of time. Experience, alertness, and ingenuity in making simple tests are valuable assets. A wiring diagram or an instruction manual may be valuable. "Case histories," such as are published in handbooks of troubles found in particular models, are helpful in some cases.

No fixed rules of procedure are possible; in fact, there are several schools of thought as to the best procedure. However, in servicing a broadcast receiver, a procedure such as the following may be pursued unless special conditions dictate otherwise:

Usually it is a waste of time to begin by turning the trimmers or other elements having to do with the alignment of the receiver. The receiver should be inspected for visible signs of electrical or mechanical defects. Next the power may be turned on while watching for smoke or odors due to overheating and listening for hum or crackles. If the rectifier has a glass envelope, the plates can be observed for overheating. Glass-envelope tubes can be observed for luminescence, indicating gas in the tubes. The tube envelopes may be felt to see if they are warm. The tubes may be checked in a

¹ E. W. HEROLD, An Analysis of the Signal-to-noise Ratio of Ultra-high-frequency Receivers, *RCA Rev.*, **6**, 302, January, 1942; D. O. NORTH, The Absolute Sensitivity of Radio Receivers, *RCA Rev.*, **6**, 332, January, 1942.

tube checker (done by many servicemen as the first step). If these attempts have not located the trouble, the customary procedure is to make dynamic tests with a signal source, starting at the loudspeaker and working toward the antenna and listening for anticipated responses in the loudspeaker. An experienced serviceman frequently can save time by identifying correct and incorrect operation through the reaction of the receiver to simple tests. For example, in some cases removal of a power-amplifier tube should result in a thump in the loudspeaker; touching the grid of a detector should result in a hum or a squawk. When the trouble has been isolated to a particular stage, voltmeter and ohmmeter tests may be practical for locating the defective element. If a table of tube-socket terminal voltages is available, troubles frequently are located by systematic checking of the various voltages.

Alignment of a receiver should not be attempted until it is in an operative condition. The frequencies at which alignment is best performed and the procedure to be followed in a particular receiver usually are specified by the manufacturer. It is a good rule to follow the manufacturer's directions. The following outline gives an idea of how a conventional superheterodyne with a single-peaked i-f amplifier may be serviced.

The i-f amplifier generally is aligned first. A modulated signal having a carrier frequency equal to the specified intermediate frequency is applied by means of appropriate connections to the signal grid of the mixer (or converter) tube. In certain cases it may be desirable to disconnect the local oscillator. The receiver should be operating with maximum gain when aligned. The AVC system sometimes is disconnected, or the test signal is kept so weak that the AVC system is ineffective. The i-f transformers are adjusted repeatedly in turn to secure maximum receiver output. In the receiver shown in Fig. 5.2 the adjustments would be made on capacitors C_{11} , C_{13} , C_{17} , C_{19} . The loudspeaker may be used as an output indicator, or an output meter may be connected at some point in the audio system, usually across the loudspeaker terminals.

Next, the signal generator is connected to the antenna and ground terminals of the receiver. The receiver tuning dial is set to a frequency near the high-frequency end of a particular tuning range, and the signal generator is adjusted to this frequency. The input voltage being kept as small as possible (readjustment being made to a weaker input signal as the alignment proceeds) the r-f and oscillator trimmer capacitors (which are not shown in Fig. 5.2 but

which are in parallel with C_3 , C_8 , C_{30}) are adjusted for maximum receiver output. Then the receiver dial is set near the low-frequency end of the tuning range under adjustment, and the signal generator also is set to this frequency. Again with the weakest possible test signal, the oscillator padder capacitor (C_{28} , Fig. 5.2) is adjusted for maximum receiver output. The procedure is repeated, perhaps several times.

For more precise alignment of single-peaked i-f amplifiers or for the alignment of a receiver having a broad double-peaked i-f frequency response, a frequency-modulated test signal may be employed and a cathode-ray oscillograph used as the output-indicating device. The patterns seen on the cathode-ray oscillograph screen not only indicate the state of alignment but also may give information concerning the shape of the response curve of the receiver and the presence of phase distortion, regeneration, oscillation, or hum.

CHAPTER XXIV

TIMING CIRCUITS

1. General Remarks.—A vast amount of recent research directed toward the practical application of cathode-ray devices has led to the development of electronic circuits especially designed to control the motion of an electronic beam and to turn it off and on according to a predetermined schedule. In some cases, the beam must trace an elaborate spiral, radial, or zigzag path upon the fluorescent screen, and the timing of all parts of this complex cycle must be correct within one-fifth of a microsecond. Modern television standards, for example, require this extreme precision, and the methods and devices originally developed for television have been adopted widely in radar practice and copied wherever a precise control for measurement of time is required.

Radio-echo measurements, originally designed to determine the height of the reflecting layers of the ionosphere, have been refined by the direct influence of television experience, until it is easily possible to deal with the weaker signals and the shorter time lags characteristic of echoes from air-mass boundaries within the earth's lower atmosphere (tropospheric echoes). As a radio wave travels 300 m/microsec, an error of $\frac{1}{2}$ microsec in timing an echo results in an error of 30 m in locating the air-mass boundary. Such precision is ample for the purpose and permits accurate timing of echoes from reflectors well within 1 km of the transmitter, as well as echoes from more distant points.

In addition to the precise timing of an individual cathode-ray beam, the control circuit must frequently provide means for accurately synchronizing a group of similar devices. In order to meet such exacting requirements the electronic network frequently becomes an elaborate device, often including dozens of tubes. In fact, television control devices incorporating more than a hundred tubes have been considered not at all exceptional.

Fortunately, these elaborate circuits may be analyzed into simple component parts, analogous to the individual machine tools that form a part of a long assembly line within a large factory.

In fact, this analogue is a particularly useful one, as the individual tubes and circuits often are strung out in a long chain of successive units. A synchronizing signal, for example, may be formed by a series of simple operations, performed by one tube after another, which clip off an unnecessary portion of the waveform here, fit on an additional corner there, square up the leading edge of a pulse, and then shave down an undesired hump. Like a factory superintendent familiar with each stage in the construction of a motor-car, the experienced operator can detect quickly the source of an error in the resultant waveform and can make the necessary correction in the individual operation that is at fault. Individual "electronic assembly lines" often lead to a final assembly line, where the carefully fabricated subassemblies are merged together by suitable electronic switchgear. Like individual bricks in a large and intricate wall, the components of the control circuit are small and simple and are charged usually with a single operation of rather elementary character. The apparent complexity is due to multiplicity of parts and tends to disappear when the circuit is reduced to block-diagram form.

Many of these individual components are familiar devices, such as amplifiers, rectifiers, cathode followers, filters, delay networks, and all sorts of sinusoidal oscillators. Other components, such as discriminators, reactance tubes, and saturable reactors, may be less familiar. However, they do have important applications in other specialized electronic circuits. In addition, there is a number of generators of nonsinusoidal waves, such as rectangular waves, saw-tooth waves, and triangular waves, together with counting circuits and other devices for frequency division and multiplication. Phase shifters, alternators, and potential dividers are included. Pulse makers of all sorts are appropriate. Finally, electronic switchgear is required for changing connections among groups of such component parts, the switching to be accomplished with maximum precision and speed.

2. Simple Relaxation Oscillator.—As an elementary example of a control circuit, examine a simple relaxation oscillator, designed to generate a saw-tooth wave for the purpose of providing a linear sweep circuit in an inexpensive cathode-ray oscilloscope.

Figure 2.1 is a familiar basic circuit, which has been analyzed in detail in Chap. VI. The fixed voltage \bar{E} applied suddenly to a circuit containing R and C in series produces an exponential rise of charge on the plates of the capacitor C . If a battery of 200 volts,

a capacitor of $0.01 \mu\text{f}$, and a resistance of 100,000 ohms are assumed, the circuit has a time constant CR of 0.001 sec. As the difference of potential between the plates of the capacitor is at all times proportional to the instantaneous charge on the plates, the potential e_c of point A rises exponentially above the ground potential, approaching the battery voltage asymptotically, Fig. 2.2.

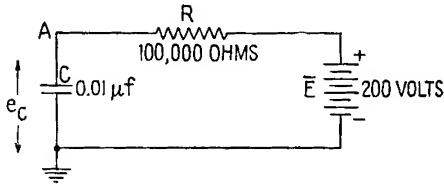


FIG. 2.1.—Basic circuit.

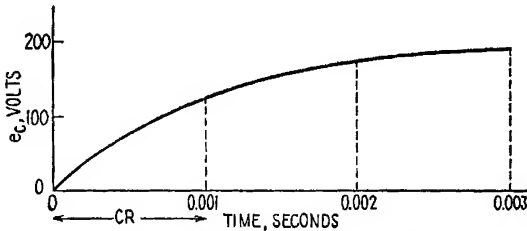


FIG. 2.2.—Exponential rise of voltage.

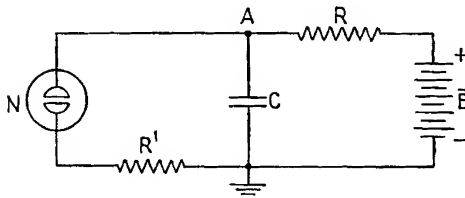


FIG. 2.3.—Simple relaxation oscillator.

For example, it would require 0.001 sec for the potential e_c to reach 126 volts and 0.002 sec for it to reach 173 volts, and the charging process would taper off exponentially if allowed to continue without interruption. Suppose, however, that across capacitor C a switch is connected which will close automatically at the instant the potential difference reaches some predetermined threshold value. In particular, this automatic switch might be a small neon bulb, Fig. 2.3.¹

¹ The effect may be demonstrated with an ordinary commercial neon bulb, commonly sold as an economical night light. The stabilizing resistor con-

The evacuated glass bulb N contains a pair of electrodes in the form of simple plates or disks and a small amount of neon gas at low pressure. Until an arc is struck, at a potential difference of about 72 volts, the tube may be regarded as an open switch, exerting no influence on the normal charging transient of the RC circuit, except by adding a few micromicrofarads to the capacitance of capacitor C .

As soon as the critical potential of the tube (say 72 volts) is attained, an arc suddenly strikes between the electrodes, resulting in a characteristic orange-red flash as the neon gas ionizes. This arc practically short-circuits the partly charged capacitor C , as the resistance R' of the arc and the external wiring is always small in comparison with the resistance R , which forms a part of the charging circuit. In consequence, the potential of point A begins to fall rapidly, the time constant $R'C$ of the discharge path being much

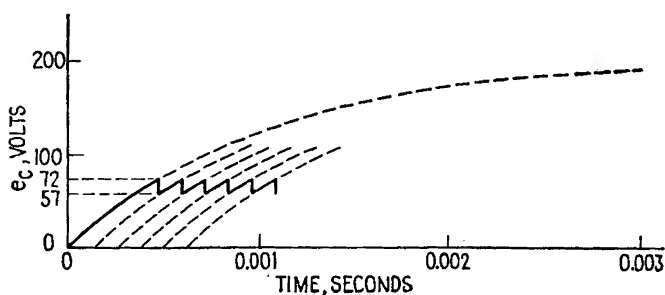


FIG. 2.4.—Production of saw-tooth wave.

smaller than the corresponding time constant RC of the charging circuit. Since the voltage necessary to *maintain* an arc is considerably less than the voltage required to *strike* the arc, the rapid discharge will continue until the potential of A has fallen to approximately 57 volts (a measured value for one particular commercial neon tube, being used merely for the sake of illustration).

At this point, the arc suddenly goes out, and the charging process is resumed. In other words, after this initial setback, the voltage again rises from 57 to 72 volts, following the appropriate portion of a simple exponential curve, precisely similar to the first charging curve, though necessarily shifted bodily to the right on the time scale. A voltage of 72 having again been reached, the arc is struck once more, and the whole cycle continues repeatedly, Fig. 2.4.

tained within the brass ferrule is not needed and may be short-circuited after drilling a small hole in the metal shell.

The output voltage, measured from point *A* to ground, is approximately a saw-tooth wave. More accurately, both charge and discharge lines are sections of exponential curves, but the range of variation of potential is so restricted that the departure from linearity is slight. To express this in other terms, the charging current changes about 11 per cent during the charging period. As a result, the potential at the mid-point of the charging period is 64.7, whereas it should be 64.5. In consequence, in a "linear" oscilloscope sweep with a total base line of 4 in. the fluorescent spot would be 0.05 in. beyond its mid-scale position at the time when it should be crossing this position.

This elementary oscillator may be improved in a number of ways, but before discussing such improvements this simple device will be used as an example to illustrate frequency adjustment, synchronization, frequency division, frequency multiplication. Such ideas may be introduced with less distraction in connection with the elementary oscillator. Once accepted, they will apply equally well to the more elaborate circuits to follow.

3. Frequency Adjustment.—The slope of the exponential charging curve, hence the period of the oscillation, may be controlled by varying the time constant of the charging circuit. In the simple oscilloscope this is advantageous since it permits a wide and continuous range of sweep frequencies, controllable in a simple manner. In examining the waveform of a 7,450-cycle alternating-voltage wave, for example, the operator may adjust his "linear" sweep circuit to 7,450 cycles, if he wishes a detailed view of 1 cycle, or may use any convenient submultiple, such as a 1,490-cycle sweep, if he wishes to exhibit 5 cycles of the alternating-voltage wave.

For convenience, the frequency adjustment often has a coarse control (range control) switch, which changes the capacitor *C* in fixed steps, and a fine control knob, which varies the resistance *R* continuously. However, there is a practical lower limit below which *R* may be reduced. By Ohm's law, it is evident that too low a resistance *R* will prevent the voltage *e_c* from ever falling below 57 volts and hence the arc will glow continuously, stalling the oscillation.

Any alteration in the voltage *E* will *also* change the slope of the charging curve, thus affecting the frequency. In practice, this control ordinarily is not needed for intentional adjustment of frequency; but it appears nevertheless as a nuisance effect, causing the sweep frequency to vary slightly and to cause a resultant side-wise drift of the alternating-voltage waveform under observation.

Automatic synchronization is necessary in order to prevent this drift.

4. Synchronization.—In electronics, “synchronization” usually denotes the injection of a periodic control voltage into an oscillator circuit for the purpose of stabilizing the frequency. When this has been accomplished successfully, the oscillator is said to be “locked in” with the control voltage. In general, the periodic

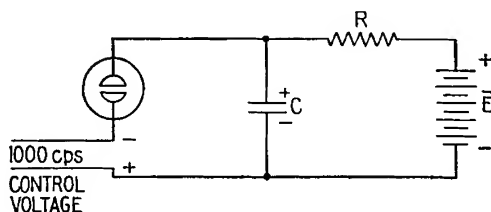


FIG. 4.1.—Introduction of control voltage.

control voltage may have any sort of waveform, and the oscillator may be sinusoidal or intentionally nonsinusoidal.

In particular, synchronization may be illustrated by connecting a relatively weak sinusoidal voltage in series with the neon lamp just discussed, as in Fig. 4.1. The alternating control voltage will be called “positive” at the time when it tends to promote the discharge. Half a cycle later it will tend to oppose the discharge and then will be called “negative.” To be specific, assume a 1,490-

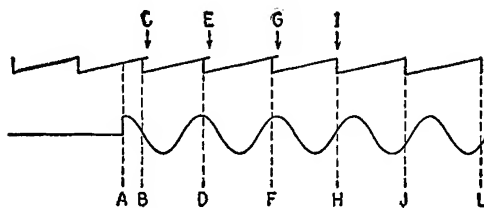


FIG. 4.2.—Synchronization process.

cycle control voltage, of $\frac{1}{2}$ volt maximum amplitude. Assume, further, that the natural frequency of the saw-tooth oscillator is slightly higher than that of the control voltage, so that eventually it will be necessary to *lengthen* each cycle by a small amount in order to attain synchronization. This may be accomplished as follows:

With the uncontrolled oscillator in operation at its own natural frequency, let the synchronizing voltage be added suddenly by closing a switch (at time A, Fig. 4.2). At the first instant, the

relative timing of the two cycles is wholly fortuitous; assume that the initial phase relationship happens to be that represented at time *A*.

If left to its own devices, the next discharge would begin at time *C*. However, in this particular instance, the initial effect of the control voltage is to *shorten* the cycle already in progress, since a weak *assisting* (positive) voltage happens to be available at the instant when the discharge is about to take place. Alone and unaided, the control voltage would be altogether too small to ionize the gas. Superimposed, however, upon the rising voltage across the capacitor, this weak assistance causes the breakdown to occur at time *B*, slightly ahead of its normal schedule.

The discharge is completed in routine fashion, and a new charging cycle begins. Being shorter than the sine-wave cycle, the saw-tooth wave during its next cycle tends to meet an earlier portion of the sine wave, and discharge would begin at time *E* if the control voltage did not interfere. This time, however, the assisting voltage is slightly larger, and in consequence the discharge actually begins at time *D*, thus further speeding the natural drift of the relative position of the two waves. On the next recurrence the discharge may take place on the leading quarter cycle of the positive crest (time *F*). At this point, the extra compression of the cycle from *G* to *F* is now becoming less important, though no expansion of the cycle has yet begun, and the shift therefore continues.

Eventually, however, the discharge will pass through condition *H* and will finally arrive at condition *J*, somewhere on the trailing side of the negative crest. With the control frequency lower than the natural frequency, this condition *J* offers the only opportunity for stable operation. For any charging cycle, terminating on this trailing quarter of the sine wave, the discharge will be retarded and the complete saw-tooth cycle stretched. For some particular phase position within this last quarter cycle the amount of stretching will be just sufficient to bring the saw-tooth cycle into exact agreement with the control cycle. Relative sliding of the two waveforms continues until this exact position is attained. When this is finally arrived at, the synchronizing transient is terminated and a stable controlled oscillation continues at the exact frequency of the external source.

The intervening transient may last for a few cycles or a few hundred cycles, depending upon the amount of enforced alteration of frequency demanded, the relative strength of the control voltage,

the relative phase of the discharge encountered at random when the control switch is thrown (at time A). It is evident that a different random choice of the initial phase relationship shown in Fig. 4.2 would result in general in a shorter transient period.

Requirement of a greater departure from the natural uncontrolled frequency would cause the discharge point on the sine wave to slide automatically to a position of higher negative voltage so that the synchronizing voltage at the instant of discharge is larger in magnitude. If the departure is carried to an extreme, the amplitude of the control voltage would have to be increased in order to maintain control. If the natural frequency and the control frequency agree, the control voltage might be reduced to a minimum. In any case, the discharge would approach the point where the sine wave crosses the zero axis.

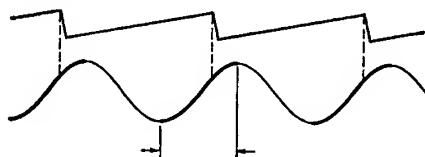


FIG. 4.3.—Possible range of phase drift without loss of synchronization.

By constructing a similar diagram it may be shown that a sawtooth wave may be *compressed*, if the natural uncontrolled frequency is too low. The discharge point slides, during the synchronizing transient, until it arrives at its one appropriate stable position, which now is somewhere on the *leading* side of the positive crest. (Though a voltage of appropriate sign and magnitude exists also on the *trailing* side of the positive crest, this position is wholly unstable, the slightest fluctuation in position being immediately augmented by the consequent change in the duration of the sawtooth cycle. On the stable side of the crest, similar fluctuations are self-correcting.)

Consequently, a change of line voltage or a drift of temperature may cause an automatic shift of phase of the discharge time, relative to the control wave, within the limits indicated in Fig. 4.3.

Hence, without changing the phase of the control wave, it is theoretically possible to exercise a limited control over the phase of a sweep circuit simply by varying the amplitude of the control voltage or the “fine control” of the natural frequency. In general, this interlocking of adjustments is undesirable; it is better to employ some form of phase shifter to slide the *control* wave forward or backward. The sawtooth generator will follow smoothly without a

perceptible transient unless the phase shift is exceptionally sudden and violent.

When precise timing is of great importance, care must be taken not to interrupt the timing wave, even for a few cycles; for the resulting drift of the oscillator will necessitate a recurrence of the synchronizing transient, and this may spoil the normal operation for several cycles after the control is resumed.

5. Oscilloscope Synchronization.—In the preceding section, a sine wave is used as a specific illustration of a control signal, but it is evidently quite unnecessary to restrict the waveform of the control voltage. Practically, any systematically recurrent cycle or pulse would be acceptable and could be used to “lock” the sweep voltage by appropriate adjustment of the sweep-oscillator circuit. Consequently, most oscilloscopes have suitable switches on the control panel, so that the operator may inject into the horizontal sweep oscillator a proportionate part of whatever recurrent voltage he happens to be examining. The signal to be examined is connected in the normal fashion to the input terminals of the amplifier, which lead to the vertical deflecting plates. Within the oscilloscope cabinet, however, a cross-connection may be established at will for the purpose of feeding a small part of the signal voltage into the “linear-time-base” saw-tooth oscillator.

6. Pulse Synchronization.—In particular, numerous applications involve the use of regularly spaced positive pulses, frequently of very short duration, as the control signal. In this case, the natural frequency of the uncontrolled oscillator should be adjusted to a value slightly lower than the desired stable rate. Hence, the positive pulse will have an opportunity to terminate the charging cycle, trimming the normal cycle slightly until the saw-tooth wave matches the control pulse precisely. If the leading edge of the positive pulse can be made nearly vertical, the pulse is said to have a steep wave front. Such a pulse will lock the saw-tooth wave so positively that it will be substantially free from slight phase fluctuations arising from line-voltage irregularities. Naturally, due care must be taken to ensure that the pulse signal itself shall show no irregularities and no interruptions.

Pulse synchronization is the standard method used in maintaining the exceedingly high precision required in television sweep circuits. Comparable refinements in the saw-tooth generators intended for such applications will be discussed later.

In principle, it would be possible to give the relaxation oscillator

a natural frequency slightly above the desired stable rate, slowing it down by a broad negative control pulse. Stability of phase would require that this pulse have a steep *trailing* edge with precise regularity of recurrence. In practice, this method has not found favor, and the positive pulse normally is employed.

7. Frequency Division.—Note that a weak control signal will be ignored by the oscillator unless and until the control signal occurs at a time when a normal discharge is nearly due. The weak signal is effective only when it occurs on top of a higher voltage eventually developed by the oscillator itself, in its own good time. Consequently, any intermediate pulses, occurring at a time when the capacitor is quite insufficiently charged, will do no harm. In fact, they should be completely ineffective in changing frequency or waveform. For example, a saw-tooth oscillator, having an uncon-

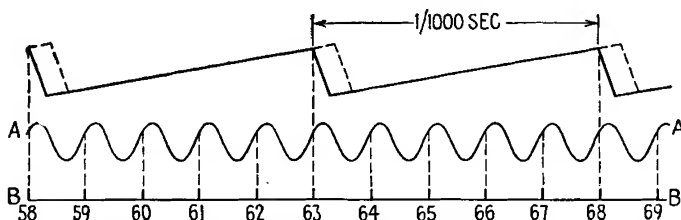


FIG. 7.1.—Frequency division.

trolled natural frequency of 950 cps, should be “locked” at precisely 1,000 cps whether it be controlled by 1,000 by 2,000 by 7,000 pulses per second. The use of a sinusoidal control voltage, of multiple frequency, is equally permissible, as the intermediate crests will be equally ineffective. The positive half cycle cannot accelerate the discharge and the negative half cycle cannot retard it until one of these repeated attempts is made at a time when the discharge is nearly ready to fire.

Figure 7.1 indicates a saw-tooth wave, stabilized at 1,000 cycles by a 5,000-cycle alternating voltage (control voltage A). Similar results would be obtained by using the 5,000-cycle pulse voltage B. In fact, the use of short, sharp pulses offers a slight advantage, as a wider engineering margin is allowed for possible drift of synchronizing-signal amplitude or slow alteration of circuit “constants.” In the figure, the pulses and cycles have been numbered in an arbitrary sequence. When the oscillator is first connected up and the control signal injected, there is ordinarily no assurance that a discharge will begin at time 58. In fact, there is initially only one chance in

five that this particular half cycle or pulse will be selected. However, a regular recurrence rate is established soon after the synchronizing signal is first injected; and having occurred at some time, here called 58, the discharge will repeat every thousandth of a second, returning at time 63, time 68, etc.

Consequently, one may generate a standard frequency, such as 50,000 cycles, by means of a crystal oscillator or equivalent device, and then obtain accurate submultiples of this master frequency by using it to control appropriate oscillators stabilized at the desired "subharmonic" frequency. Correctly constructed "locking" circuits, being assumed, the percentage accuracy of any low frequencies thus obtained should be as good as that of the standard frequency from which they all are derived.

However, there is danger that erratic action may develop in the frequency divider if one attempts to use a frequency ratio larger

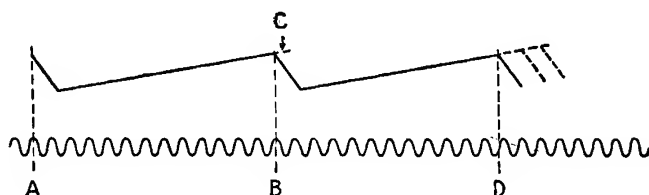


FIG. 7.2.—Erratic frequency division.

than about 10. In fact, where even a momentary error may cause serious inconvenience, conservative engineers often prefer to keep the ratio as low as 7. The reason for such limitations is indicated in Fig. 7.2. Assume that a 2,184-cycle control signal is available and that the operator intends to divide by 13, thus producing a 168-cycle saw-tooth wave. Having begun one discharge at time A, the oscillator, of its own accord, would repeat the discharge at time C. With the aid of the control voltage, however, the saw-tooth cycle is shortened and the discharge correctly tripped at time B, after an interval of exactly 13 cycles of the control wave. On the next occasion, however, a minor upward fluctuation of the charging voltage may slightly accelerate the charging process. The voltage across the capacitor is then so near the breakdown voltage that it may accidentally flash over at the twelfth cycle instead of waiting properly for the thirteenth. This particular sweep, therefore, is shortened excessively. This momentary error shifts the entire saw-tooth pattern ahead in phase and may necessitate manual readjustment.

Evidently a steady upward drift of the charging voltage soon would cause the circuit to produce an erroneous stable frequency of 182 cps. A downward drift would result in a frequency of 156 cps. Similar confusion would result from a fluctuation or a drift in the amplitude of the control wave. In both cases, the trouble is due to the fact that the twelfth, thirteenth, and fourteenth crests are too near together, as too many control cycles have been crowded into one cycle of the relaxation oscillator. By using a more conservative ratio a wider engineering margin is allowed, and such rigid control of supply voltage, control amplitude, and circuit temperature need not be required.

8. Frequency Multiplication.—In common with all other non-sinusoidal generators, any relaxation oscillator may be used for frequency multiplication. The straight sides and sharp corners of

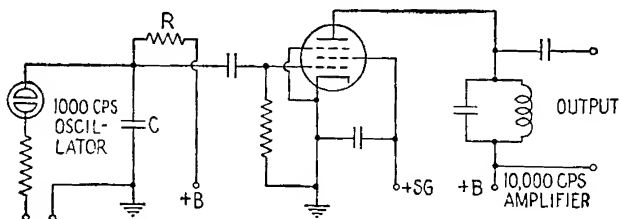


FIG. 8.1.—Circuit for frequency multiplication.

the voltage wave indicate that the output wave is rich in harmonics of high order and considerable intensity. Therefore, multiplication may be accomplished by means of any sharply resonant circuit capable of selecting the desired harmonic and attenuating all others. To prevent reaction upon the oscillator, which might spoil the waveform at its source, the high- Q parallel-resonant circuit is frequently placed in the plate circuit of a buffer amplifier, the control grid being driven by the oscillator, Fig. 8.1.

A ten-to-one frequency step-up is easily attainable in one stage, and considerably higher ratios sometimes are employed. Frequently, a decimal series of frequencies may be desired. In any case, the multiplication may be continued in successive stages (cascade multiplication).

Persons who have had little practical experience with the Fourier series representation (*i.e.*, the inevitable existence of harmonics in any recurrent but nonsinusoidal wave) are frequently puzzled by the implied exactness of the multiplication and ask whether a slight detuning of the resonant circuit would not cause the

output frequency to deviate gradually from its nominal value. The following physical explanation therefore is offered:

Assume that a saw-tooth oscillator is securely locked in with a precise 1,000-cycle control frequency. What happens when this oscillator drives a buffer amplifier that has a sharply resonant output circuit, tuned only approximately to 10,000 cps?

The high- Q resonant circuit may be compared to a massive pendulum driven by repeated blows from an external source. The breakdown of the neon tube or equivalent tripping device in the relaxation oscillator suddenly delivers energy to the tuned circuit, simulating shock excitation. One single discharge would produce in the tuned circuit a transient of small amplitude but long duration. Most of the energy of the individual transient would be found at the natural frequency of the resonant circuit, which would depend upon the exact setting of the tuning capacitor. However, for a random setting, the repeated and overlapping transients produced by periodic impacts would interfere destructively with one another, yielding a negligible net response. Only when the interval between impacts is an exact multiple of the period of oscillation will the initial transient be strongly reinforced and built up. Consequently, on gradually tuning a sharply resonant circuit, steep response peaks should be obtained at the exact frequencies of the harmonics of the nonsinusoidal driving voltage.

The frequency step-up obtained per stage is limited only by the strength of the harmonics in the driving wave and by the "resolving power" of the resonant circuit. A sharp-cornered wave is rich in harmonics and is useful in this application. As stated above, ratios as high as 10 are obtainable rather easily. Unlike the frequency-division circuits there is little likelihood that the multiplication circuits will suddenly produce an incorrect frequency as a result of a minor voltage or temperature drift. Instead, the output will fall off progressively, indicating the need for correction.

9. Frequency Addition and Subtraction.—Note that frequencies may be added and subtracted, *i.e.*, sum and difference frequencies may be obtained, by the use of a nonlinear "mixer," which is described in connection with modulation and frequency conversion in Chaps. XX and XXIII. Note also that the *linear* "mixer," to be described later, is intended for an entirely different purpose.

10. Thyatron Sweep Circuits.—The general principles of synchronization, frequency division and frequency multiplication have been discussed above in terms of an elementary form of relaxation

oscillator, almost as simple in its behavior as the typical doorbell, which it somewhat resembles. In considering various improvements in the oscillator itself, it is to be noted that, while these improvements lead to circuits which are more elaborate, the same general principles apply.

By employing a hot cathode in the discharge tube and inserting a grid between cathode and anode, the elementary circuit of Fig. 2.3 may be changed, to the thyatron¹ circuit of Fig. 10.1. As in the gaseous diode, the tube operates as a single-pole single-throw switch, open during the charging period and automatically closed for the discharge. As before, the gas within the tube suddenly ionizes when the voltage across C attains sufficient magnitude. However, any increase of the adjustable negative grid bias produces a proportionate increase in the breakdown voltage of the tube.

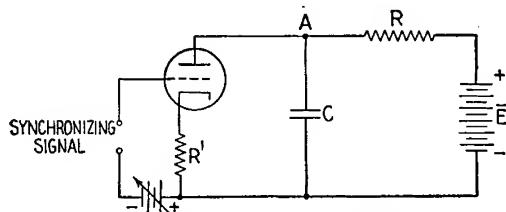


FIG. 10.1.—Relaxation oscillator using gaseous triode.

For moderate negative bias the grid loses control as soon as the arc is established. Hence the arc continues until the available plate voltage drops to a value insufficient to maintain ionization. The net effect of the grid-bias control therefore is to extend and control the amplitude range of the saw-tooth voltage, simultaneously altering the natural frequency.

Any periodic voltage introduced into the cathode lead permits synchronization (together with multiplication or division if desired) precisely as in the diode circuit previously discussed. But, by injecting the control voltage into the grid circuit, equivalent results are obtained with a considerably weaker synchronizing signal. The gain of sensitivity is about 10 to 1 in a typical case. Note particularly that the circuit continues to function as a self-excited saw-tooth oscillator when the control signal is suppressed entirely. In normal operation the control voltage is a very minor ripple, unable to "fire" the discharge unless the plate potential has already climbed nearly to the breakdown point. A positive crest lowers

¹ Chap. XVI, Sec. 5.

the discharge threshold momentarily and slightly accelerates the discharge. A negative crest has the opposite effect. Variations of grid potential are ignored by the tube until the capacitor is charged sufficiently.

11. improvement of Linearity.—On taking advantage of the increased range of potential variation obtainable from the three-electrode tube, Fig. 10.1, it becomes increasingly apparent that the voltage across C increases exponentially with time. In general, an accurately linear increase would be preferred, as a larger fraction of the available source voltage could be used.

The following mechanical analogue may be suggestive: Allow water to flow through a small pipe, projecting nearly to the bottom of a cylindrical barrel, the pipe being fed from a reservoir of fixed height. If the reservoir is high, the water rises in the barrel at a

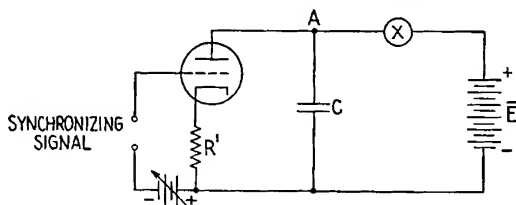


FIG. 11.1.—Improvement of linearity by use of current limiter.

substantially constant rate, since the rate of flow through the pipe is nearly independent of the rising head of water within the barrel. Without requiring a high reservoir the same result could be secured by designing a valve that would limit the flow to a fixed number of quarts per minute, regardless of the water pressure exerted. It could be agreed that the difference in level would never be allowed to fall below some convenient lower limit (say 5 ft above the barrel top).

In the electrical case, several different devices are available for this purpose. The symbol X in Fig. 11.1 represents any current-limiting device, used in place of the charging resistance R , Fig. 10.1, to determine the constant rate at which the charging current trickles into capacitor C .

For example, X might represent a simple diode, with a tungsten filament, the space current showing a sharp upper limit, dependent only upon cathode temperature. So long as the battery voltage always exceeds the voltage across C by a margin sufficient to utilize all the electrons that the cathode will release, any variations in the potential of point A can have no effect upon the constant charging

rate. The potential therefore rises linearly, discharge finally occurring just before the margin conditions are violated.

More commonly, X represents a pentode, connected as indicated in Fig. 11.2. With the designated bias potentials the current through the two-terminal device is over a suitably restricted range, independent of the voltage across these terminals.

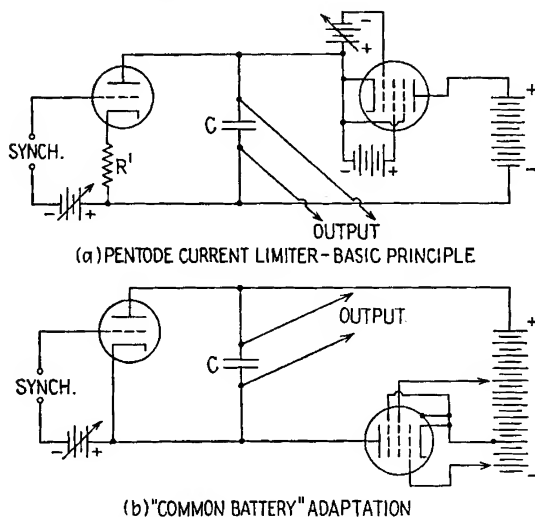


FIG. 11.2.—(a) Relaxation oscillator with pentode current limiter; (b) "common battery" adaptation. In practice, rectifiers with potential dividers replace the batteries.

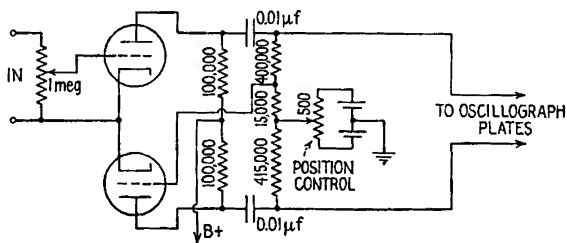


FIG. 11.3.—Phase inverter.

Other forms of limiter are available. It should not be inferred, however, that limiters are universally employed even in precision work. Frequently the designer prefers to generate the saw-tooth wave at such a relatively low level that no correction is needed. He then introduces sufficient amplification before applying the saw-tooth wave to the oscillograph. In the amplifier the sweep level is conveniently adjustable; and, by means of a phase-reversing cir-

cuit, equal and opposite voltages may be applied to the two horizontal plates of the oscillograph, thus maintaining balance with respect to ground. A typical amplifier for this purpose is shown in Fig. 11.3. Cathode-bias resistors and capacitors, appropriate to the tubes in use, may be inserted as usual for adjusting the grid bias.

12. Thyatron Pulse Circuits.—In previous diagrams it is implied that the low resistance R' includes merely the arc resistance plus the necessary circuit wiring.

Suppose, however, that a few hundred ohms are deliberately added in the cathode circuit of Fig. 10.1, R' now standing for this external resistor. On discharge, the voltage of capacitor C is suddenly applied to this low resistor in series with the still lower resistance of the arc. A sudden pulse of voltage, similar in shape to the current wave that produces it, appears across the resistor R' . Upon subtracting the small internal drop in the ionized tube, the initial amplitude of the voltage pulse approaches the potential attained by capacitor C . Like the discharge current, the pulse voltage falls nearly exponentially at first, then drops more rapidly as the arc starts to go out. As a result, this pulse has steep sides and a short duration, in order of magnitude equal to the time constant $R'C$.

In some applications this pulse may be desired *in addition to* the saw-tooth wave produced by the same circuit. For example, on suitable inversion of polarity, it might be used to turn off the cathode-ray beam while the oscilloscope retrace is in progress. In this case, it would be called a "blanking pulse."

In other cases, the thyatron circuit may be set up for the sole purpose of making a short, sharp pulse of easily controllable duration. For example, pulse circuits have been in use for many years in radio-echo experiments conducted at Cruft Laboratory. In this application the pulse modulates a continuous-wave radio transmitter, applying radio-frequency voltage to the antenna for controllable periods ranging from a few microseconds up to about 200 microsec. It may do this by applying plate voltage to an oscillator or to an r-f amplifier tube, the tube having no other plate supply than the pulse; or the pulse may remove a grid-blocking voltage that normally interrupts the r-f circuit at any convenient stage in the amplifier. Because of the small fraction of the time occupied by the actual transmission, a large power output may be obtained from a small tube.

In such applications, where there is no need for the saw-tooth wave and the interest is in the discharge transient, it is possible to economize by simplifying the charging circuit. Accordingly, Fig. 10.1 may be converted into Fig. 12.1.

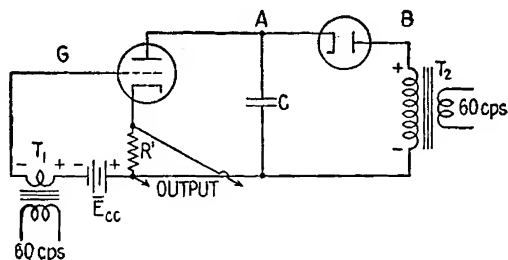


FIG. 12.1.—Thyatron pulse circuit without d-c source.

This represents a saving of apparatus and power. In practical applications of Fig. 10.1, the voltage \bar{E} is usually obtained from a "power pack." In Fig. 12.1 this power pack has been reduced to a simple transformer and half-wave rectifier. No filter is necessary or desirable. The high resistance R has been removed entirely, with

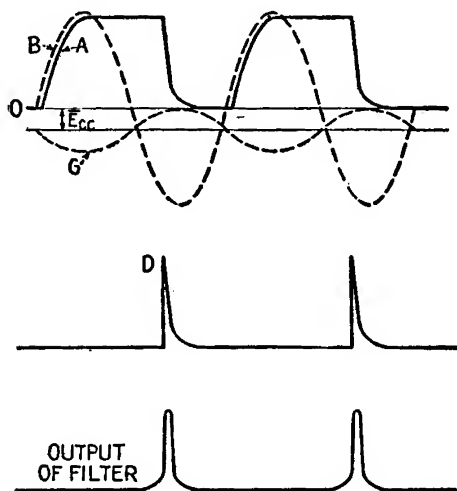


FIG. 12.2.—Performance of thyatron pulse circuit of Fig. 12.1.

consequent improvement of efficiency. In place of the saw-tooth wave the cycle shown in Fig. 12.2 is obtained.

Driven by transformer T_2 , the potential of point B varies sinusoidally as indicated in Fig. 12.2. At the crest of the positive cycle, B reaches a level slightly higher above ground than the peak

value of the pulse (which is to be produced later). For example, assume 500 volts as the amplitude of the sine wave at B . As the total resistance in the charging circuit is now low, the time lag is negligible and the potential of A rises practically in step with B , remaining slightly lower than B because of a small drop in voltage across the rectifier tube. Assume that point A has been lifted 485 volts above ground at the end of the first quarter cycle. Because of the low time constant of the charging circuit, transient effects are negligible, for the present purpose, during the charging period. Hence all trace of the exponential growth has been lost, together with the approximate linearity that resulted from it. Instead, the voltage across capacitor C follows the applied sinusoidal potential without significant lag.

As the B curve starts downward, at the beginning of the second quarter cycle, the A curve ceases to follow. Electrons, drained away from the upper plate of the capacitor during the first quarter cycle, cannot be restored, for there is no route for them to follow. The rectifier is a one-way gate, which now bars return by the original path. The alternative route through the thyatron is not yet available, as the plate voltage is insufficient to produce ionization. Hence the positive potential of the upper plate of the capacitor is suddenly "frozen," just below the crest value of the applied sine wave. This potential would remain constant forever after, if other parts of the circuit did not eventually intervene.

This intervention usually takes place during the third quarter cycle. An alternating voltage, obtained from the same 60-cycle source, is applied through transformer T_1 to the grid of the thyatron. Merely by selecting the proper polarity of the secondary leads, the potential of G , Fig. 12.1, may be made to vary in opposite phase to the potential of A . Therefore, during the charging period and the remainder of the first half cycle, the small voltage of T_1 augments the bias battery, keeping the grid so negative that the discharge cannot begin. During the third quarter cycle, however, T_1 offers increasing *opposition* to the bias battery, thus facilitating the discharge. When the net negative voltage is brought sufficiently near zero, the steady plate potential is suddenly sufficient to "fire" the tube, producing the characteristic exponential pulse shown at D , Fig. 12.2. If desired, a phase shifter may be used to supply the control voltage, thus permitting the pulse to be moved through wide limits and locked at any position in the second, third, or fourth quarter of the cycle determined by voltage A .

Note that the voltage supplied by T_1 is no longer a “synchronizing voltage” in the sense previously employed. The discharge is now completely under the control of the voltage supplied by T_1 , the circuit having no significant natural frequency. The altered function of this voltage may be emphasized by calling it a “control voltage.”

As an added refinement, the output pulse obtained from the circuit of Fig. 12.1 may be passed through a simple filter, as indicated in Fig. 12.3, before being used to modulate a transmitter. This is to remove unnecessary side humps from the frequency spectrum of the emitted r-f pulse and hence to minimize interference with other radio services. A “probability curve” is an ideal pulse envelope for general pulse-modulation purposes. Unlike other

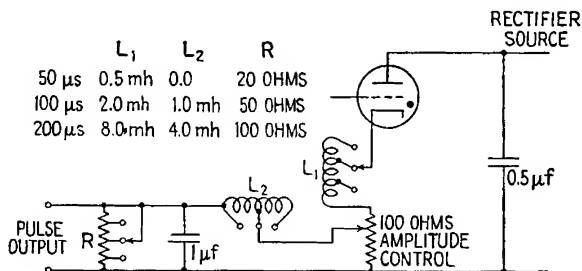


FIG. 12.3.—Pulse-shaping filter.

functions, this pulse has a frequency spectrum that closely resembles the waveform of the modulation envelope. In fact, this frequency spectrum is *also* a “probability curve” (Chap. XVIII). Therefore, the side bands fall off smoothly and symmetrically from the carrier wave of the modulated transmitter. This ideal can be approximated very well in practice without serious difficulty. Direct pulse modulation of an *oscillator* maximizes interference, for it tends to produce a violent frequency sweep together with the rise in amplitude of the wave.

13. High-vacuum Discharge Tubes.—To return now to the discharge tube, which is to be used for producing a saw-tooth wave, would it be feasible to substitute a high-vacuum tube in place of the thyatron? If this could be done, we could hope to obtain the advantage of somewhat longer life with greater stability of the characteristics. Gaseous tubes are of great value in those electronic devices which cannot be made to function without them, but the high-vacuum tube is generally preferred when a choice

becomes possible. This moderate prejudice is largely due to the fact that the gas pressure varies with room temperature and may drift upward or downward as the tube ages. Apparently *some* gas is constantly being evolved from the tube elements or adsorbed on the glass walls and metal surfaces. In consequence, the operating characteristics are less stable than in a comparable high-vacuum device.

Fundamentally, the discharge tube is required to operate as a single-pole single-throw switch, capable of changing very suddenly from an open circuit to a comparatively low resistance element in the network. In the thyatron this sudden change is provided by the striking of the arc, which takes place very quickly and violently, even though the synchronizing signal on the grid may be changing smoothly through a small voltage range. If this same synchronizing signal were applied to the grid of a "hard" tube, the tube would act as an ordinary amplifier, the instantaneous resistance changing smoothly and continuously in accordance with the grid modulation. This would not produce the sudden switching action that is required. If, however, a series of short, sharp, very powerful positive pulses could be delivered to the grid of the high-vacuum tube, the circuit then would imitate the action of the thyatron as a switching device.

Normally, the tube would be biased below cutoff by a steady negative potential. Each positive pulse would overcome this negative bias and would immediately carry the instantaneous grid voltage so far into the positive region that the tube would be able to carry a heavy instantaneous current with comparatively little plate voltage. The pulses would have to be delivered by a sturdy and powerful generator, capable of maintaining the positive pulse in spite of the relatively heavy instantaneous grid current that inevitably would result. Steepness of the leading edge and the trailing edge of the pulse would be essential for proper operation. The actual duration of the pulse also would be important, for by using a sufficiently short period the discharge of the capacitor could be stopped before completion.

Note that the grid of the high-vacuum tube retains control at all times. This is in contrast to the thyatron, where the grid normally loses control after the arc has been struck, the termination of the discharge awaiting the necessary drop in plate voltage, as this voltage must be low enough to allow the arc to go out.

If a high-vacuum discharge tube is used, the discharge-tube circuit ceases to be a relaxation oscillator. It does not execute a

cycle of its own if the grid-voltage control is withdrawn. Hence the grid voltage has been elevated in importance from the role of "synchronizing signal" to that of "control signal." At the same time, the specifications in regard to this signal and the source that must produce it have been greatly stiffened. The substitution of a hard tube becomes possible because these stiffer requirements may be met without serious difficulty.

14. Blocking Oscillator.—Obviously, there would be little or no advantage secured by eliminating the gaseous tube from the discharge-tube circuit if it were found necessary to reinstall it as an essential component of the pulser circuit, which drives the discharge tube. Hence, although thyatron circuits are capable of delivering short and powerful pulses, in this particular application an equivalent hard-tube circuit is to be preferred. The "blocking oscillator" is one example of such a device.

Other examples, such as trigger circuits and multivibrators, will be discussed later.

In the blocking oscillator, a familiar fault of oscillator circuits which is frequently encountered as a nuisance effect, is deliberately

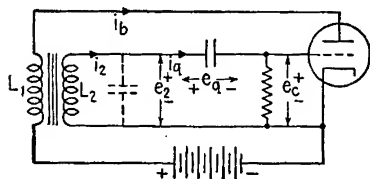


FIG. 14.1.—Blocking oscillator.

exaggerated and converted into a useful source of regularly spaced transients. Intermittent oscillation, resulting from improper choice of oscillator grid leak and grid capacitor, is described in Chap. XV. The interruption of oscillation eventually occurs as a result of a negative grid bias that grows too large and is held too long. In the so-called blocking oscillator this interruption is made to occur very quickly, so that the "oscillation" ceases before completing a single cycle.

The circuit resembles the familiar tuned-grid oscillator circuit. However, the regenerative feedback is abnormally strong, close coupling from plate circuit to grid circuit being obtained by means of a step-up transformer with an iron core. The usual tank-circuit tuning capacitor is omitted, though the secondary of the transformer necessarily has distributed capacitance, Fig. 14.1, which determines a natural frequency.

In descriptions of the blocking oscillator, it is often implied that the pulse emitted by the oscillator is the first positive half cycle of what would ordinarily be a continuous oscillation, the remainder of the wavetrain being immediately suppressed, as the grid becomes

biased beyond cutoff. Careful examination of waveforms in the laboratory reveals that the action is not quite so simple as this elementary explanation implies. As an example of the utility of the cathode-ray oscilloscope in quickly unraveling an apparently intricate circuit problem, these waveforms will be discussed in some detail.

On viewing the pulses displayed on an oscilloscope screen, several apparent anomalies appear. In place of the rounded and approximately sinusoidal pulse implied by the conventional explanation, it is often apparent that the plate-current pulse, Fig. 14.2, can be extremely angular, having steep sides, sharp corners, a downward-slanting top, and appreciable width. As expected, the secondary voltage e_2 rises very rapidly as the primary current i_b

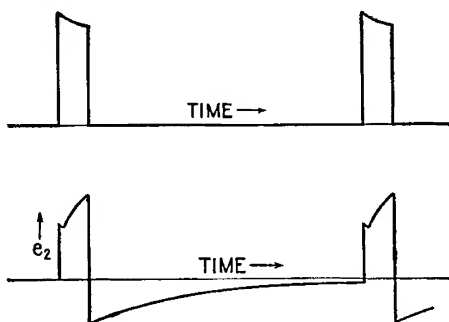


FIG. 14.2.—Pulse waveform.

rises. However, after the i_b curve has attained its peak value and definitely is slanting downward, the e_2 curve continues to rise for an appreciable period. Finally, the slope of the falling i_b curve changes abruptly, and a rapid drop of e_2 ensues.

The necessary explanation may be phrased in two different ways, both of which are useful. Begin by considering the behavior of the magnetic field in the transformer core. In analyzing the transient interactions of vacuum-tube devices in general, it is frequently desirable to provide a clear and definite starting point for the discussion. This can be done best by assuming that the plate power supply has been turned off for a long period and is suddenly restored.

At zero time, therefore, the plate current suddenly begins. The initial rise of plate current in the primary coil induces a voltage in the secondary coil that is applied to the grid. Being connected in regenerative fashion, the induced grid voltage is positive and consequently accelerates the plate-current flow. This, in turn, induces

a higher secondary voltage, which produces a still higher current, etc. The cumulative process would continue indefinitely were it not for the ceiling imposed by cathode emission or, more commonly, by Ohm's law. Even if the plate of the tube were short-circuited directly to the cathode, the current could not rise above the limiting value determined by the power-supply voltage and the resistance of the external plate circuit.

Having quickly attained its ceiling value, the plate current suddenly stops rising. At first glance, one might expect that this would immediately withdraw the source of the positive grid voltage and that in consequence the plate current would instantly drop. However, as mentioned above, the positive secondary voltage continues to rise, and the plate current remains high for a definite interval. (During this interval the current is not constant but decreases at a moderate rate.)

The explanation hinges on the fact that the magnetic flux in the transformer core does not at once attain the equilibrium value, which would be computed from a knowledge of the magnetomotive force due to primary current *alone*. Instead, by Lenz's law, one must expect a *counter* magnetomotive force as long as secondary current continues to flow. As this secondary current dies away, the primary current remaining large, the magnetic flux continues to increase. An increasing proportion of the high primary magnetomotive force becomes available to overcome the reluctance of the core. Finally, the demagnetizing force vanishes, and the flux is momentarily in equilibrium with the primary current alone.

The secondary current is made up of two parts, an internal portion due to distributed capacitance of secondary winding, and a more important external current that flows through the tube and through the grid capacitor. This external current is mainly due to the electrons rapidly acquired by the positive grid. Initially the grid capacitor was uncharged, and hence the entire positive secondary voltage e_2 was available on the grid. As the electron current continues to flow, an increasing proportion of the large positive voltage e_2 is expended in maintaining the ever larger charge on the grid capacitor. During this period the electrons are accumulating on the grid capacitor much faster than the grid resistor can allow them to leak away.

The grid voltage e_g is that *fraction* of the very large positive voltage e_2 which *remains* after subtracting the capacitor voltage drop e_q , Fig. 14.1. Though this *fraction* decreases rapidly from its

initial value of unity, this does *not* result in an equally rapid fall of the *actual* grid voltage e_c . Instead, the continued rise of the secondary voltage e_2 supplies most of the growing capacitor voltage e_q , with a relatively minor decrease in the initially large positive voltage e_c . However, the slight resultant decline in e_c does cause and explain the slight drop in i_b noted as the short charging transient progresses.

The charging current of the grid capacitor does not vary exponentially, since the source voltage e_2 is not constant and the grid-to-cathode resistance of the tube is nonlinear. Experimentally, for common circuit adjustments, the charging current falls almost linearly, intersecting the zero axis abruptly, Fig. 14.3. This abrupt termination of the charging transient is desirable. As the flux

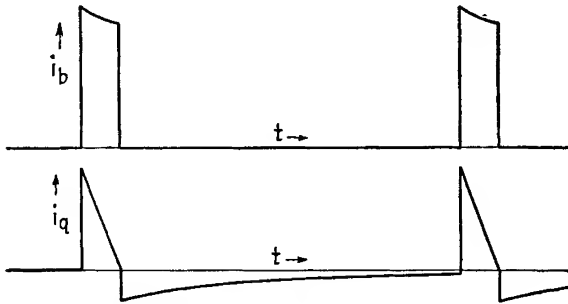


FIG. 14.3.—Primary and secondary currents.

momentarily attains equilibrium with i_b and therefore stops increasing, the high positive voltage e_2 is suddenly withdrawn. For an instant this leaves only the reverse voltage e_q in the external grid circuit. This fact, by itself, would be sufficient to cause the grid to assume a negative potential considerably beyond cutoff and to *retain* a negative potential until the charge has had time to leak away. This action is reinforced considerably by the coincident collapse of the *magnetic* flux, which supplies an *additional* reverse voltage, driving the grid still further into the negative region. The collapse of flux is retarded by the secondary current of the discharge transient, and the *induced* negative voltage therefore also persists for an extended time. The whole time scale of the discharge is sluggish, because the discharge current, driven by e_q and the reversed e_2 , now has to pass through the high-resistance grid-leak resistor R . Hence the tube retains a negative bias beyond cutoff, for a period which depends largely upon the time constant

CR. This period may readily be made large in comparison with the duration of the charging transient.

The same explanation, just presented in terms of magnetomotive force and flux, also may be expressed in terms of mutual inductance M and self-inductance L_2 . Such terms may be considered as an abbreviated notation, or "shorthand," that offers the advantage of mathematical compactness but permits some loss of contact with the underlying physical phenomena. Occasionally it is desirable

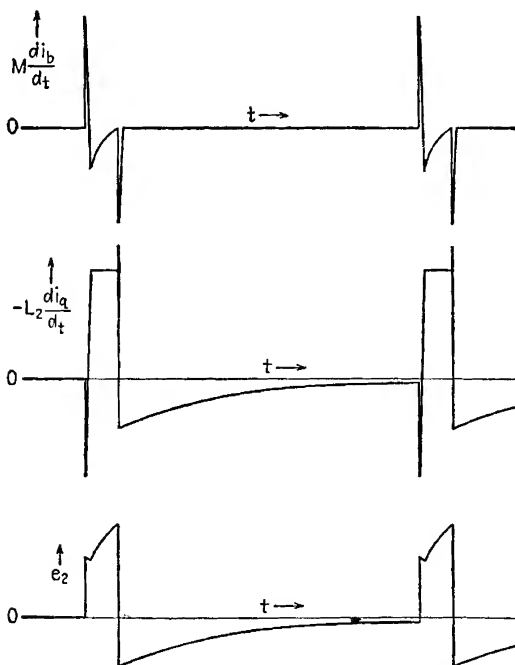


FIG. 14.4.—Analysis of secondary voltage wave.

to dig underneath this convenient superstructure, as has been attempted in the preceding paragraphs.

In order to apply the inductance concepts, begin with the observed oscillograms of i_b and i_q , Fig. 14.3.

The voltage e_2 may be obtained from

$$e_2 = M \frac{di_b}{dt} - L_2 \frac{di_2}{dt}$$

In this expression, i_2 properly should include the combined currents necessary to charge all elements of the distributed capacitance of

the coil and wiring. In practice, however, these internal currents appear to be relatively small in comparison with i_q , which may be observed and measured conveniently in the external circuit. Representation of i_2 by i_q leads to an adequate interpretation of the voltage waveforms actually observed.

Hence

$$e_2 = M \frac{di_b}{dt} - L_2 \frac{di_q}{dt}$$

Therefore the waveform of e_2 may be predicted graphically by determining the slopes of the curves of Fig. 14.3 and then summing the two components of e_2 as indicated in Fig. 14.4.

The computed wave e_2 agrees with the corresponding measured wave e_2 , Fig. 14.2. The drop in potential across C also may be

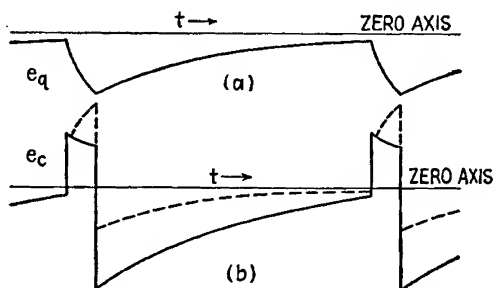


FIG. 14.5.—Analysis of grid pulse.

obtained experimentally and is shown in Fig. 14.5a. Upon deducting this graphically from e_2 , Fig. 14.5b results, which represents the predicted wave e_c . This computation is verified also by experiment.

15. Saw-tooth Voltage System.—By the term “system” is meant an assembly of component devices, all working together to produce a desired result.

An incidental advantage of the blocking oscillator lies in the fact that it may be coupled to a “discharge tube” in a particularly simple and effective fashion, Fig. 15.1.

The grids and cathodes of the two triodes may be joined directly, no other grid connection being necessary in the discharge tube. The positive side of the e_c cycle, Fig. 14.5b, agrees very well with the specifications for the positive control pulse (short, sharp, powerful) that is to actuate the switching tube. The slight variation in amplitude during the existence of the positive pulse merely affects the exact shape of the retrace line, which is of little consequence so long as the retrace is concluded speedily.

The waveform of the negative side of the e_c cycle is immaterial until the potential finally rises to the cutoff value. Preferably, the two tubes should have similar characteristics, so that the discharge tube is released from cutoff just as the corresponding release occurs in the pulser tube. The halves of a twin triode are often used for this reason, forming a compact and economical system. By producing a sharp positive pulse of considerable amplitude, while carrying the load represented by its own grid current, the pulser has already demonstrated its ability to "stand up" under load. The addition of the second grid is equivalent merely to an alteration in the design details of the original triode and does not alter the general nature of the cycle just discussed.

As usual, the desired saw-tooth voltage wave appears directly at the terminals of the capacitor C , Fig. 15.1. Connection to the

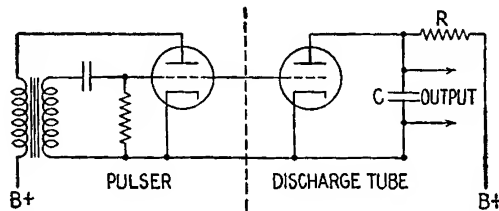


FIG. 15.1.—A saw-tooth voltage system.

oscilloscope or other load circuit is made preferably through a "buffer" amplifier, which is another element in the complete system.

16. Saw-tooth Current Wave.—As explained in more detail in Chap. XII, Sec. 6, the sweep coils of a magnetically deflected oscilloscope require a peculiar alternating voltage wave, which is neither saw-tooth nor rectangular, but a definite mixture of the two. This voltage must be supplied in order that the *current* wave shall be of simple saw-tooth form.

Fortunately, a very slight modification of Fig. 15.1 will yield an excellent approximation to the desired voltage wave. In Fig. 16.1, a high resistance R limits the charging current as usual; resistance R' stands for a few ohms of resistance in the external wiring, but resistance R'' is a new low-resistance element inserted in series with C . Unlike Fig. 11.2a, both R'' and C are now to be included between output terminals. Again the output voltage will be delivered through a linear amplifier. At the frequencies concerned, one may consider that the amplifier draws no current from the output leads shunted across the CR'' coupling element.

The amplifier merely transmits and amplifies the voltage wave, delivering it through a power-amplifier stage, which then drives the deflecting coil as a load, Fig. 16.2.

Assuming that a positive grid-voltage control pulse has just terminated, consider the subsequent variations in the potential of point *A*, Fig. 16.1. At the beginning of the new charging cycle, point *A* is positive with respect to ground, the potential of *A* being made up of two parts. The larger part is due to the fact that the

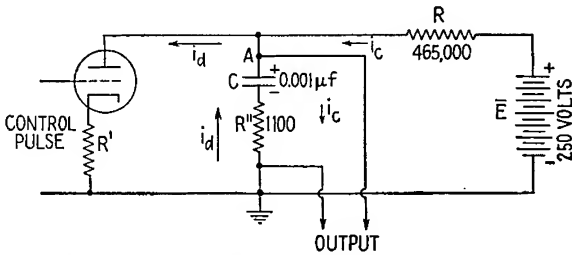


FIG. 16.1.—Voltage source for magnetic deflection in a cathode-ray oscillograph.

capacitor *C* was not completely discharged by the grid pulse just ended. The discharge was of limited duration and was suddenly cut off while the capacitor still retained a considerable charge. To be specific, assume that the potential of the upper plate of capacitor *C* is 40 volts above that of the lower plate as the new charging cycle begins.

In addition, a small potential difference exists across the coupling resistor *R''*, the upper terminal of this resistor being positive

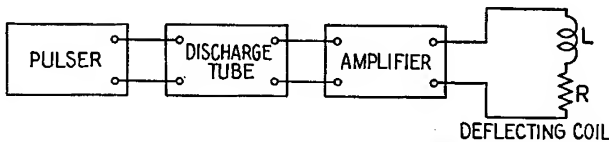


FIG. 16.2.—Block diagram of circuit for producing magnetic deflection.

with respect to ground. Assume this potential difference to be approximately 0.5 volt. It is due to the charging current *i_c*, say 450 microamperes, flowing through the coupling resistor toward the lower terminal. Upon adding this small positive voltage to the 40 volts across the capacitor *C*, the total output voltage, measured at point *A*, is 40.5 volts.

During the first 1/18,000 sec following removal of the positive pulse the potential difference between the plates of a 0.001-μf capacitor *C* would rise linearly from 40 to 65 volts if the charging

current were to remain constant. Actually, the current falls off by about 10 per cent because of the increasing opposition offered by the capacitor C . If the resulting slight departure from linearity is unacceptable, it may be reduced by increasing the applied voltage \bar{E} and the resistor R in proportion. To the degree of approximation that the voltage rise may be regarded as linear, the $i_c R''$ voltage may be regarded as 0.5 volt during the entire charging period.

Suddenly a new control pulse arrives on the grid. The direction of the current through C and R'' reverses instantly at the moment that the low-resistance discharge path first becomes available. Assume that the initial value of discharge current is about 20 times as large as the charging current, say 9 milliamperes. This would correspond to a tube resistance of approximately 6,100 ohms (including R' , but not including R'') while the tube is in the conducting state. As a result of the sudden reversal of current in R'' , the potential of point A immediately drops 10.5 volts. Of this amount, 0.5 volt drop may be ascribed to the loss of the small charging current, while the remaining 10 volts is due to the new discharge current i_d , flowing upward from the lower terminal of the coupling resistor R'' .

Even though the resistance in series with C is comparatively low, the difference of potential across C cannot change instantly. Hence the initial drop in the output voltage measured at A is entirely due to the reversal of current through R'' . This vertical *drop* of 10.5 volts is immediately followed by a steep downward *slope*, as the increment of about 25 volts, slowly gained by the capacitor during charge, is completely lost during the swift discharge.

If it could be assumed that the discharge current remained constant, this additional loss of potential would occur linearly and would carry the potential of point A down to 30 volts. Just before termination of discharge, the lower plate of capacitor C would be 10 volts below ground potential, the upper plate being 30 volts above ground. Sudden removal of positive voltage from the grid (as the control pulse snaps off) instantly lifts the lower plate of C from -10 volts to $+0.5$ volt as the charging process is resumed. This brings the potential of A back to its starting point at 40.5 volts. The entire cycle then repeats as indicated in Fig. 16.3. It should be understood that the numerical values calculated are introduced solely by way of explanation. Widely different values may be used successfully.

The voltage waveform obtained is a combination of a rectangular

wave with a saw-tooth wave and agrees precisely with the specifications (Chap. XII, Sec. 6) for driving a saw-tooth current wave in a load circuit containing self-inductance and resistance. This waveform may be described as "a peak riding upon a saw tooth" and sometimes is called a "trapezoidal" voltage wave. The relative proportions of peak and saw tooth may be altered conveniently by making R'' continuously variable. The operator adjusts the peaking resistor R'' until satisfied with the linearity of the resultant current sweep. In practice, it is not necessary that the *discharge* current remain approximately constant. If the tube resistance were unchanged during the time occupied by the control pulse, the current would follow a portion of an exponential decay curve. Actually, the control pulse often does *not* have an accurate flat top, and

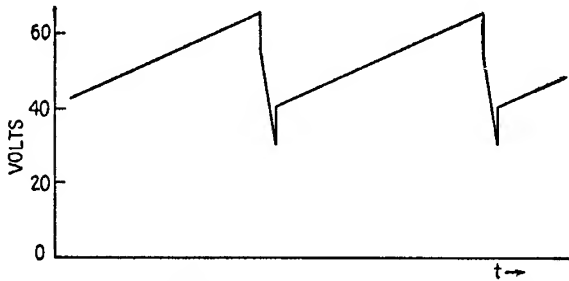


FIG. 16.3.—Rectangular voltage plus saw-tooth voltage.

the discharge current follows no simple mathematical function. However, this departure from the simple theory as presented merely alters the exact shape of the strong negative pulse added to the saw-tooth wave, Fig. 16.3. This alteration, in turn, produces a departure from linearity in the retrace portion of the saw-tooth current curve. Considerable nonlinearity of retrace is ordinarily acceptable as long as the sweep portion of the saw-tooth cycle remains approximately linear and the retrace is completed on schedule.

Therefore, the purpose of peaking resistor R'' is to add to the voltage saw-tooth wave a short, sharp negative peak of controllable amplitude. The exact shape of this peak is of secondary concern. The mere existence of the peak provides the necessary assistance in reversing very quickly the current in the deflection coils of the oscillograph at the end of the linear sweep, thus preventing interference with the next cycle.

17. Clippers.—In order to understand the function of a "clipper," let it be assumed that on a long strip of paper an accurate

curve has been constructed which represents instantaneous values of a current or a voltage, plotted as a function of time. The time axis runs lengthwise along the strip. The waveform thus constructed may have any shape that can be produced in an electrical network or picked up by an antenna. It is not even necessary that the variation shall be periodic. With a pair of scissors, this strip of paper is sliced into two (or more) ribbons of fixed width, the cutting line running exactly parallel to the time axis. If preferred, the ribbons may be of quite dissimilar width, as the cut may be made at any level selected on the positive or the negative side of the zero axis.

After slicing, one of the ribbons ordinarily is retained and used. The upper strip or the lower strip may be discarded at will. In some cases, two separated portions of the original wave are usefully employed but are routed to different destinations after being cut apart.

This perfect "shearing" action represents an ideal, which the practical clipper approaches rather closely. Obviously, an ordinary half-wave rectifier may be regarded as a special application of a clipper. In this case, the "electronic shears" cut directly along the zero axis itself. Different types of rectifiers differ in the degree to which they approximate the ideal behavior suggested. Any of the rectifiers may be used as a component in the generalized clipper circuits represented in Fig. 17.1.

In Fig. 17.1, a sinusoidal input-voltage wave is used arbitrarily as an illustration, but it will be apparent that no restrictions need be placed upon the waveform. The input is applied at terminals 1-2. Resistor R has a resistance that is high in comparison with the low internal resistance presented by the rectifier unit while conducting. Voltage output terminals 3-4 often feed into a buffer amplifier, which draws a negligible input current in the frequency range involved. In any event, the impedance of the load circuit should be high in comparison with R . The battery \bar{E} stands for any source of d-c bias potential. In most cases, this will be some fraction of a voltage derived from a rectifier and filter, used in common by various elements of the control circuit. Line 2-4 is often grounded, for convenience elsewhere in the circuit.

In Fig. 17.1a the rectifier cannot conduct until the potential of point 1, measured with respect to line 2-4, becomes larger than +10 volts. For all negative values on the input-voltage wave and for all positive values smaller than 10 volts, the plate of the rectifier

is at a lower potential than the cathode. Under this condition, electrons cannot cross the tube, and the shunt circuit might just as well be absent. As the load circuit 3-4 draws little or no current from the source, the series resistor R has little or no effect during this portion of the cycle. Hence the output potential at point 3 accurately reproduces all the variations of input potential, until the input potential rises above the bias voltage. At this point, the rectifier suddenly ceases to be an open circuit and becomes instead a conducting element. In the allowed direction of conduction the

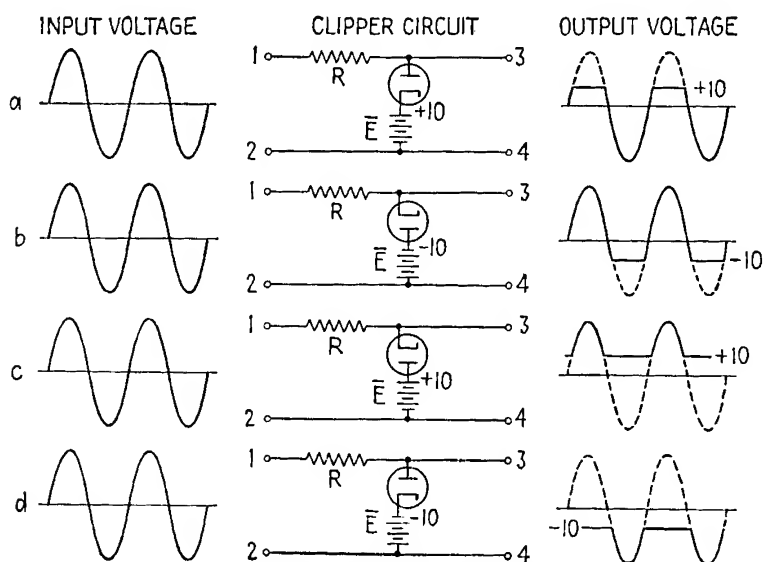


FIG. 17.1.—Diode clipper circuits.

resistance of this element is low in comparison with the resistance of the series resistor R . Hence the output voltage suddenly ceases to follow any further increase that may be applied to the input terminals. Instead, such further increments in input voltage are absorbed almost entirely by the resistor R , producing a current through R and through the bias source \bar{E} , without appreciable voltage rise across the rectifier. The shunt current flows in such a direction as to carry energy from the input terminals 1-2 toward the bias battery or equivalent d-c source, part of this energy being dissipated en route in the resistor R . The remainder is used in charging the bias battery \bar{E} , or reducing the total load on the equivalent power pack. Under ideal conditions, the voltage output

terminals 3-4 require no energy flow. The current in the shunt circuit ceases as soon as the input voltage drops to the bias value. As a result, the dotted portion of the output-voltage wave, Fig. 17.1*a*, is completely suppressed, and the solid line represents the waveform obtainable at terminals 3-4.

By reversing the rectifier or the battery or both, the three other circuits shown in Fig. 17.1 are obtained. The output waveforms are exhibited in the figure. All four of these combinations are of equal interest and value, and the reader should satisfy himself, in detail, that all the output waveforms presented in Fig. 17.1 are correct. Discussions similar to the above apply equally well to Figs. 17.1*b*, *c*, *d*.

With the general purpose of clipping in mind, it becomes evident that an individual experimenter could devise readily a number of other electronic devices yielding waveforms similar to those pre-

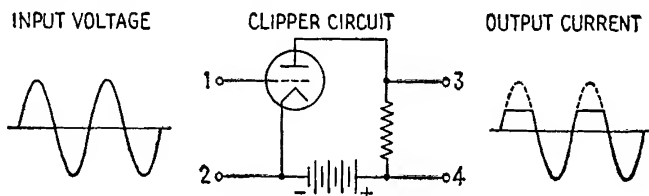


FIG. 17.2.—Saturated-triode clipper circuit.

sented in Fig. 17.1. For example, the dotted curves, previously discarded, could be clipped off and used if an output voltage were measured across R . Two or more shunt rectifiers, similar to Figs. 17.1*a*, *b*, could be used in conjunction with the same series resistor R , in order to remove the top and bottom of a wave at the same time. An example of this sort appears in Fig. 18.2. Various combinations of series rectifiers would be appropriate. Clipping and amplification often could be combined in triode, tetrode, and pentode circuits.

A few additional examples will suffice. For example, one may assume that the signal to be clipped is applied directly to the grid of a tungsten-filament triode, the filament temperature being adjusted so that saturation effects appear within the available range of operation, Fig. 17.2.

In this case the upswing of plate current terminates abruptly when all electrons emitted by the cathode are acquired by the plate circuit. Below this limiting current the grid retains control, and a linear response may be obtained. On striking the ceiling imposed

by filament-temperature limitation, the plate-current wave suddenly flattens out, and a similar intentional distortion appears in the output voltage. The resultant clipping action may be varied by altering the grid bias, the amplitude of the input wave, the filament temperature.

By using an input wave of sufficient amplitude, the current wave can be confined by a floor as well as by a ceiling. The instantaneous plate current cannot be less than zero. By selecting a tube in which cutoff occurs abruptly, this limitation may also be made to occur without serious departure from linearity in the transmitted portion of the cycle. For this purpose the triode should *not* have a remote-cutoff (variable-mu) characteristic. Though it is theoretically possible to cut off the top and the bottom of a wave with a single triode, this economy is not followed widely in practice, as it results

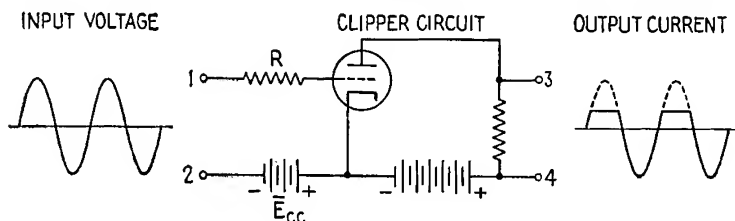


FIG. 17.3.—Triode clipper circuit in which the grid acts as a diode clipper.

in a too critical juggling of wave amplitude and bias potential. Naturally, the use of the floor type of limitation (lower bend of the characteristic curve) is not dependent upon use of a tungsten-filament tube.

It is frequently desirable to imitate the ceiling type of plate-current limitation, typified by the saturable triode, without actually having to employ this rather old-fashioned device, seldom conveniently available. One expedient is to use the circuit of Fig. 17.3, with an ordinary oxide-cathode triode, showing no actual saturation effects within the operating range. In effect, this is a repetition of the diode circuit of Fig. 17.1a, incorporated as a component part of the triode diagram. Grid current starts to flow as soon as the net grid-to-cathode potential exceeds zero, and effectively prevents the grid voltage from rising any higher. The resistance of the grid resistor R is large in comparison with the grid-to-cathode internal resistance of the tube. Because of grid-circuit rectification, the voltage actually appearing on the grid is clipped by the external resistor, no distortion of the plate-circuit

characteristic being required. As usual, the clipping level is determined by the bias battery.

As the function of the resistor R is to produce a large voltage drop when a small current suddenly flows, it is evident that a high-reactance element can be used as a substitute. For example, in Fig. 17.4 a series capacitor C is substituted for R . The capacitance of C is so small that the grid current charges it very quickly, thus keeping the grid potential from ever rising appreciably above the net cathode potential. A low-resistance grid leak R' allows this charge to escape quickly when the source voltage starts to decrease.

By comparing this circuit with other electronic circuits employing a grid-leak and capacitor combination it is evident that circuit configuration *alone* is seldom a reliable indication of the purpose of a particular circuit. In studying an actual piece of equipment one

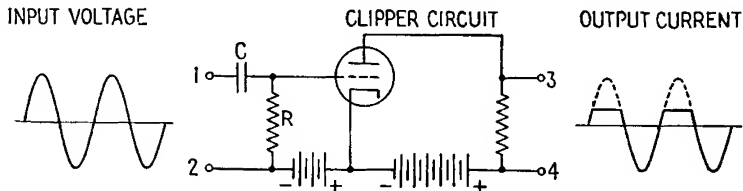


FIG. 17.4.—Clipper circuit using capacitance as a series element.

must learn by experience to give adequate consideration to circuit constants, and especially to time constants. In a normal oscillator circuit, for example, the grid capacitor usually has a negligible effect upon a *single* alternating-voltage cycle, the capacitor voltage changing very slowly in accordance with desired slow adjustments of the average amplitude. The capacitor provides an adjustable d-c bias potential only. In certain detector circuits the same configuration appears, the capacitor voltage following the modulation cycle, though not the carrier-wave cycle. In the blocking oscillator, the series capacitor responds much more rapidly and may reduce, by 50 per cent or more, the positive-pulse voltage induced in the transformer secondary and transmitted through the blocking capacitor to the grid. In Fig. 17.4 the capacitor C might be referred to also as a *blocking* capacitor, but it responds still more rapidly, being able to charge and discharge so quickly as to suppress almost entirely the net positive voltage, without appreciable distortion of the remainder of the cycle. All these divergent functions are provided by the same basic grid-leak and grid-capacitor wiring plan, the important distinctions occurring largely as a result of

alterations in the RC time constant, considered with respect to the period of the alternating-voltage cycle. Still another function, the production of sharp pulses, is discussed in Sec. 21.

As a final example of the "ceiling" type of plate-current limitation, note that pentode tubes may be made to show a sharply defined limiting action, the plate current varying linearly with control-grid voltage as long as the instantaneous grid voltage stays below a definite margin. Above this critical point the plate current remains practically constant, regardless of further increases in instantaneous grid potential. The location of the critical point depends upon the exact values of all electrode potentials and may be determined readily from the static characteristics of a particular tube. Hence a pentode amplifier with an apparently normal circuit configuration may serve as an excellent clipper if the bias potentials are selected with this end in view. The same effect frequently occurs accidentally.

As sharp cutoff effects may also be obtained with appropriate pentodes and tetrodes, the "floor" type of clipper likewise may appear in the configuration of a normal amplifier stage, careful attention to bias and amplitude values being necessary in order to discover that the designer or manufacturer has incorporated the clipping action at this stage in a circuit chain. A special application of pentode clippers is shown in Fig. 18.3.

18. Rectangular-wave Generators.—Rectangular-wave generators (also referred to as square-wave generators) are designed to



FIG. 18.1.—Square waveform.

produce an angular waveform such as that indicated in Fig. 18.1. Shall this waveform be referred to as an alternating voltage or current of 10 units amplitude, shall it be referred to as a series of pulses of 20 units positive amplitude, or of 20 units negative amplitude? The terminology depends upon whether a base line is inserted at the axis of symmetry, or at the bottom, or at the top of the figure. In some applications this is a matter of importance. In other cases it is a matter of indifference, the load circuit responding only to the variations in potential without regard to possible bias values. The term "square wave" refers only to the "squareness" of the corners and does not imply any particular relationship between amplitude and duration.

Such waveforms are utilized for a wide variety of purposes. They are sometimes used in performing standard tests upon amplifiers and other circuit elements. In this case defects in harmonic phase or amplitude response appear as characteristic deviations from the simple rectangular input. As discussed later, they are used in operating electronic switches, throwing two or more

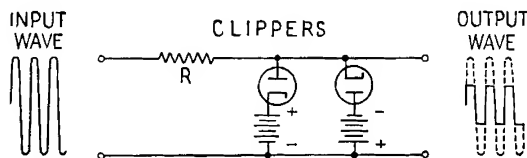


FIG. 18.2.—Elementary rectangular-wave source.

patterns on a cathode-ray screen in rapid succession, or making periodic changes in a vacuum-tube network. They are often developed as a preliminary step in the production of short, sharp pulses, the pulses being generated in a network driven by a rectangular alternating voltage.

In some cases the required rectangular voltage is generated directly by trigger circuits, having an appropriate intermittent cycle. Such trigger circuits, including multivibrators, will be discussed in Secs. 27 and 28.

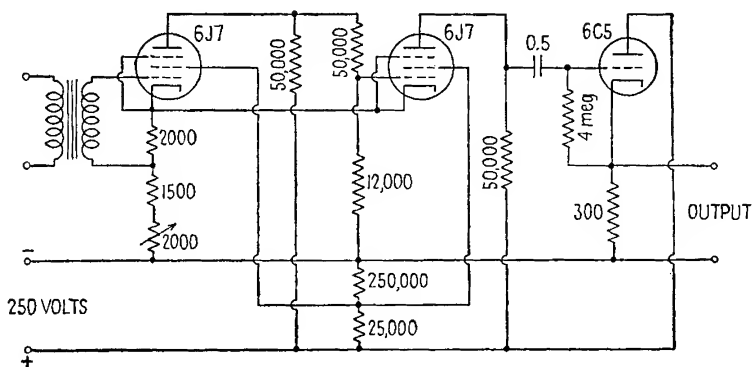


FIG. 18.3.—Square-wave clippers using pentode tubes.

In other cases the application permits an approximation, the rectangular wave being obtained by generating a sine wave of appropriate frequency and then clipping off the positive and negative crests. Where necessary, the approximation can be made exceedingly good by using a sinusoidal input voltage of large amplitude and clipping both sides close to the zero axis. If desired, the

approximation may be attained in successive stages, the output of the first pair of clippers being amplified linearly and then run through a second pair of clippers. This cascade action avoids the use of excessively high input voltages, otherwise necessary in order to obtain an equally good rectangular wave.

Figure 18.2 illustrates an elementary rectangular-wave source, based upon the diode clippers of Fig. 17.1. Figure 18.3 shows a practical laboratory instrument using pentode tubes as clippers.

19. Damping Tubes.—When pulse currents are driven through large inductances, as in transformer circuits, oscillograph deflection coils, and relay windings, occasionally the distributed capacitance and wiring capacitance of the circuit cause an undesired backswing and may lead to a damped train of oscillations. The circuit could be made aperiodic by appropriate network design, but this often would involve resistance loadings considered excessively large for the application in mind. By using a rectifier that conducts only during the backswing, the reverse voltage may be highly damped without affecting the main operation. In a sense this is merely a special case of clipping, but the problem is sufficiently important to merit special mention. The situation may be presented in greater detail by quoting an example.

Assume that an available thyatron pulse generator, of ample current rating, can deliver a pulse of 2,000 volts amplitude, repeated 60 times per second. The operator would like to use this periodic pulse generator as the only power source in the plate circuit of an otherwise conventional Hartley oscillator circuit. The oscillator and associated antenna are to emit radio-frequency waves for a small fraction of each successive period, the remainder of the period being available for the reception of echoes from the ionosphere or troposphere. Taking advantage of the small fraction of the time actually used in transmission, a low-power transmitting tube often may be used at the highest voltage that its insulation will withstand. The operator wishes to step up the voltage of the plate supply well above the available 2,000 volts but cannot raise the voltage at its source. Can he employ a transformer successfully between the pulses and the oscillator?

Given sufficient time, money, and patience, he may build or purchase a special pulse transformer to do the job. This will require special attention to reduction of distributed capacitance and special care to the minimization of core losses at high harmonic frequencies. Meanwhile, a standard 60-cycle transformer, of

appropriate voltage ratio, is immediately available on the laboratory shelf. Some wastage of energy and some distortion of pulse form would be freely permitted in order to save time. What other attendant disadvantages would be encountered if the standard transformer were employed?

On driving a conventional transformer, with a unidirectional pulse applied to the primary, distributed capacitance causes an oscillation as indicated in Fig. 19.1.

At least as a temporary expedient, the first positive half cycle of the oscillation would be wholly acceptable as a voltage pulse of adequate amplitude and duration. No direct objection need be raised in this instance to the backswing that immediately follows it, since the oscillator, serving as a load, can respond only to positive plate potential. However, the second positive half cycle, and all

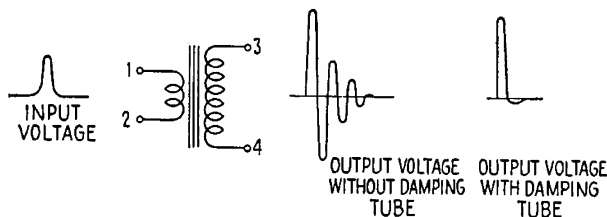


FIG. 19.1.—Transformer with a pulse input to the primary.

the positive half-cycles to follow, are sure to be confused with the echo pattern. A “fake” echo of this sort, originating in a pulsed transmitter or receiver, is a familiar nuisance in ionospheric research laboratories and is introduced to initiates as “Benny.”

By damping the backswing sufficiently, the remainder of the wavetrain may also be eliminated, including all the spurious echoes. This follows because, after the expiration of the primary pulse, the remaining transient is entirely due to *stored* energy. If this can be dissipated harmlessly during the backswing, it is obviously no longer available for producing supplementary positive half cycles. In dealing exclusively with stored energy the damping tube may be distinguished from the general run of clipper tubes, previously discussed. Figure 19.2 presents a wiring diagram showing how a damping tube could be attached to the transformer circuit discussed above.

During the first positive half cycle of the transformer transient the oscillator starts in normal fashion and has time to emit several hundred or several thousand r-f cycles before the plate supply is

withdrawn. During this period the oscillator draws power continuously from the transformer and also causes some wastage of power in the transformer windings. However, it acts as a medium-impedance load, rather than as a short circuit, and a considerable amount of energy remains stored in the transformer fields when the oscillator is turned off. Normally, this would have to be dissipated largely by the additional periods of undesired operation of the oscillator. In contrast, the addition of the damping tube practically short-circuits the transformer during the reverse half cycle, without interfering in any way with the initial pulse. The stored energy is wasted harmlessly in the internal resistance of the transformer and in the low resistance of the damping tube. Such a tube may be described as a "one-way short circuit."

The negative damping current being several times as large as the positive load current, the resistance of the transformer circuit

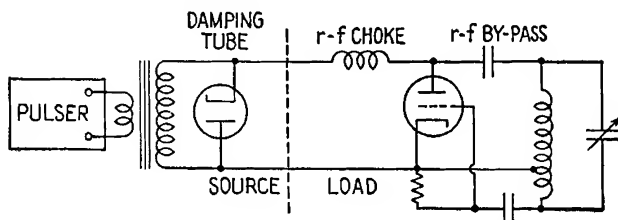


FIG. 19.2.—Pulsed oscillator.

may be deliberately increased by an external resistor of a few hundred ohms without serious interference with the oscillator circuit. This possible economy decreases the current rating demanded of the damping tube and may extend its useful life.

20. Clamping Tubes.—In Sec. 18 reference is made to the insertion of a base line, or zero line, from which instantaneous voltages or currents are to be measured. Sometimes this is merely for convenience in description. In other circumstances the location of the zero level is of considerable physical importance, the load circuit responding to the steady component, as well as the alternating component, of the input signal. In a normal amplifier chain, which has a low-frequency cutoff (at 15 cps, for example) this steady component is lost or altered at every stage in the amplifier chain, since it is not passed through any coupling transformer or through any coupling capacitor. Where transformers are employed, this defect may be remedied immediately by inserting a suitable bias battery, or equivalent d-c source, in series with the secondary side of

the circuit. This is a common practice in transformer-coupled oscillograph amplifiers, the insertion and control of the appropriate X and Y steady components determining the location of the pattern upon the fluorescent screen. However, where coupling capacitors are employed, this simple remedy does not necessarily suffice, since transient disturbances often cause slow drifts in the steady component of the output, with resultant annoyance or complete failure of the system. As its name implies, the "clamping tube" is a tool designed to "clamp" the axis of the output signal quickly and securely to the desired zero level or to some definite positive or negative level, fixed with respect to a base line. Frequently the device has been called a "d-c restorer." However, the latter name does not do full justice to its capabilities, since the clamping level on the output side of the coupled circuit may be designed to be quite different from the steady component of the input signal.

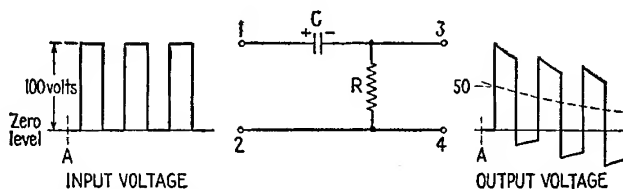


Fig. 20.1.—Slow drift of output level without clamping tube.

Figure 20.1 shows a particular example of the defect that is to be cured. Assume, for example, that the input signal consists of a rectangular voltage wave, superimposed upon a steady voltage which is equal in magnitude to the 50-volt amplitude of the alternating voltage.

The input signal is applied suddenly at time A , the capacitor C being initially uncharged. Since C is designed to be a coupling capacitor, the time constant CR will be large in comparison with the period of the applied alternating voltage. Hence the sudden jumps of $+100$ and -100 volts are faithfully reproduced at the output terminals 3-4, the charge on capacitor C being unable to change appreciably during such sudden transitions. However, as the signal continues for a number of cycles, the steady component of the input signal gradually builds up a steady charge upon the plates of C , this charge starting from zero and approaching 50 volts asymptotically. Hence the axis of symmetry of the output wave will drift sluggishly down to zero, having started at $+50$ volts. Eventually, therefore, the steady component of the input signal is

lost, though it remains, as a transient voltage, *long enough* to cause an exceedingly annoying drift of a cathode-ray oscillograph pattern or other output device. The final level attained may be shifted by insertion of a bias battery, but the slow transient persists and may be highly detrimental. A remedy is suggested by Fig. 20.2.

Insertion of the diode, Fig. 20.2, practically short-circuits the high resistance R during the first half cycle of the input wave. Instead of charging slowly to a potential of 50 volts, capacitor C very rapidly charges to the full 100 volts, acquiring practically the entire charge before the first half cycle is well advanced. Hence the sharp initial rise of potential, at output terminal 3, is followed by an immediate return to the zero axis. During the remainder of the first half cycle the output voltage remains zero. At the end of the first half cycle the voltage at point 1 suddenly drops by 100 volts, thus carrying the potential of point 3 downward by an equal

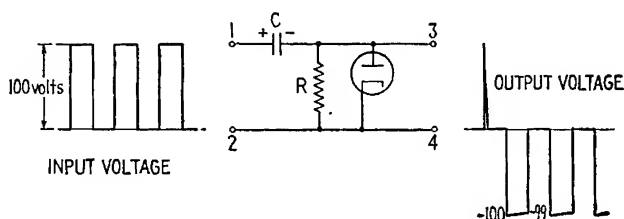


FIG. 20.2.—Rapid attainment of steady output level with clamping tube.

amount, to a level 100 volts below zero. Since the diode is now inoperative, the capacitor C would remain fully charged were it not for a very small discharge current through the high resistance R . Since the time constant CR is large in comparison with the period T of the alternating-voltage wave, this results in a very slight rise in potential during the second half cycle (say from -100 to -99 volts). At the beginning of the next half cycle, the 100-volt jump in input potential suddenly lifts the output potential to a point slightly above zero (say 1 volt positive), whence it immediately returns to zero as the diode momentarily conducts. Thereafter, the output cycle continues to be an acceptable reproduction of the alternating component of the input cycle, the *top* of the alternating-voltage wave being clamped to the zero axis.

Parenthetically, one may note that this automatic “self-biasing” action is inherent in the familiar multistage amplifier employing grid-leak and capacitor coupling. In this case the “diode” is the grid-cathode portion of each driven amplifier tube, the negative

grid bias being adjusted quickly and automatically so that the grid potential never crosses the zero axis by any significant amount. The rectangular wave of Fig. 20.2 is used merely by way of illustration. The clamping action is not dependent upon the shape of the individual cycle, though very rapid decreases of *amplitude* will not be followed immediately by the corresponding decrease of effective negative bias.

In Fig. 20.2 one may note that the clamping action has resulted in a 50-volt *negative* bias, attained immediately. Hence, by adding an equal and opposite *positive* bias, obtained from a battery or equivalent source, one may secure *without delay* the zero-bias condition that is approached so slowly in Fig. 20.1.

One way of accomplishing this is illustrated in Fig. 20.3. Here the positive bias is inserted in series with the diode, which is there-

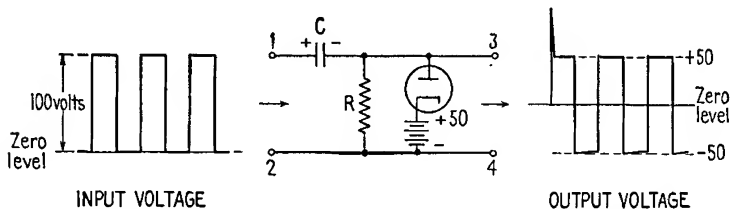


FIG. 20.3.—Attainment of zero-bias condition without delay. Top of wave clamped at +50-volt level.

fore inoperative until point 3 rises above the +50-volt level. As a result of the first violent surge of current through the diode, the capacitor C assumes a 50-volt charge, which it retains thereafter. Variation of the battery voltage permits location of the clamping level at any value from 0 to +100 volts. By reversing the battery, the top of the wave may be clamped at any desired negative level. Values of cathode bias higher than +100 volts would make the diode inoperative. (However, if the nature of the circuit permits insertion of additional bias directly in output lead 4 or lead 3, the displacement of the axis may be extended indefinitely.) Adoption of the +100-volt cathode bias would result in true “d-c restorer” action, the output wave having the same zero level as the input wave.

Reversal of the diode, Fig. 20.4, results in clamping the *bottom* of the alternating-voltage wave at an output level determined by the bias battery or equivalent bias source. In the position indicated, the bias is immediately effective in clamping the wave if the plate is held at any potential within the range of the input signal or

at any positive potential above this range. In the illustrative example, negative plate potentials between 0 and -50 volts would be partly effective, as the downward drift shown in Fig. 20.1 would be halted before completion. However, immediate stabilization could be had by inserting the negative bias directly in the output lead, Fig. 20.5.

In certain applications, "synchronous clamping" is desired. This term implies that the cathode-ray spot, for example, is to be

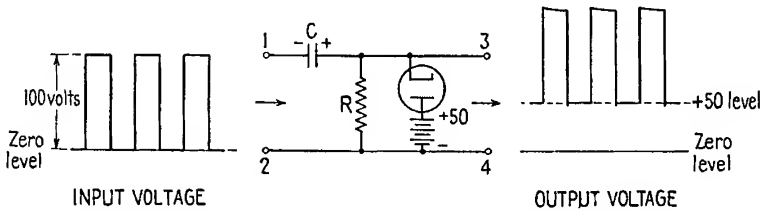


FIG. 20.4.—Bottom of wave clamped at +50-volt level.

brought exactly to a prescribed starting line (voltage level) at the *beginning* of each cycle but that the clamping action is then to be suddenly released, permitting full freedom of voltage variation during the greater part of the cycle. At the end of each cycle, clamping is to be restored. This requirement is easily met by adding a control grid between the cathode and anode of the clamping tube. Normal grid potentials do not interfere with the clamping action described above. However, sudden application of a suf-

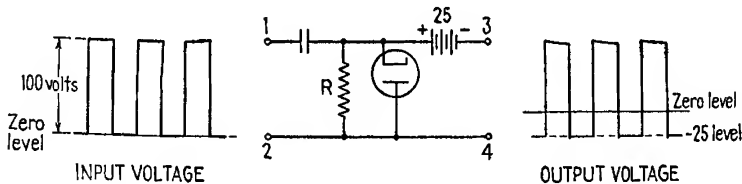


FIG. 20.5.—Bottom of wave clamped at -25-volt level.

ficiently negative grid potential will prevent current flow through the clamping tube, thus effectively removing it from the circuit. Such synchronous control requires a rectangular "keying signal" similar to that described in Sec. 24, *Keyers, or Electronic Switches*, and need not be discussed further at this point.

21. "Differentiators," or "Peakers."—In Sec. 18, reference is made to the production of short, sharp pulses by means of a network driven by a rectangular voltage wave. The simplest and most common form of network used for this purpose is illustrated in

Fig. 21.1. A capacitor and resistor are connected in series across the voltage source, the output voltage being measured across the resistor. The circuit may be regarded as an elementary high-pass filter.

The behavior of this elementary circuit is studied in Chap. VI. So far as the output voltage is concerned, the excitation of an RC transient by means of a rectangular voltage wave is equivalent to a switching process in which a battery is repeatedly introduced into

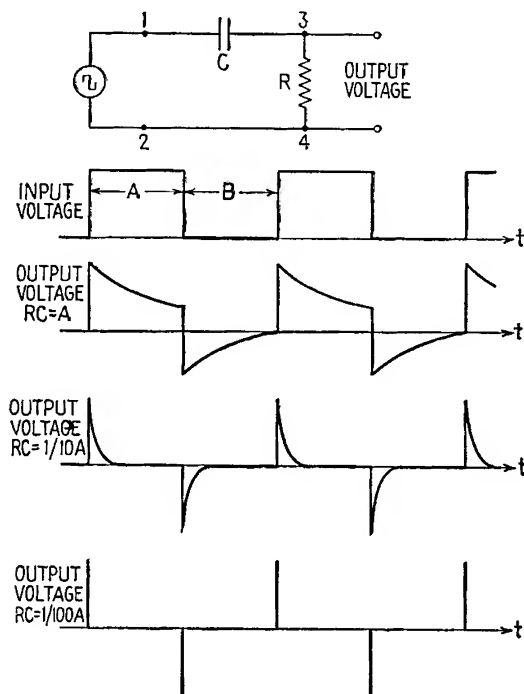


FIG. 21.1.—“Differentiator” with various time constants.

the circuit and repeatedly suppressed, without interrupting the path available for current flow. As usual, the charge and discharge currents begin suddenly and die away exponentially, the rapidity of the exponential decay being determined by the time constant RC . The impedance of the load circuit is assumed to be very high, unless it can be included in the resistance R . The output voltage is at all times proportional to the current flowing in the resistance R . The initial amplitude of each output pulse is exactly equal to the total change of voltage occurring suddenly at the input terminals. It

is assumed that the input wave has accurately vertical sides and square corners, the amplitude of the output pulse is independent of C and R . Hence, as indicated in Fig. 21.1, the exponential output pulse can be made thinner and thinner (in time duration and in total charge conveyed) as the time constant CR is progressively decreased.

Finally, the output pulse, plotted by an oscillograph on the original time scale, becomes so thin that its actual exponential character is no longer obvious. Instead, it has the appearance of a vertical pulse of negligible duration, the rise and fall being practically coincident on a time scale appropriate to the original input wave. Even in this limiting case, the pulse retains a definite finite amplitude and hence cannot be referred to correctly as the mathematical derivative of the rectangular voltage wave. Nevertheless, the general resemblance of the pulse to the corresponding derivative is rather striking, particularly as the appropriate reversals of direction are obtained at the correct positions. In any event, this similarity has led to extensive engineering use of the word "differentiator" as a suggestive name for the circuit indicated in Fig. 21.1. Recently an alternative name, "peaker," has been proposed as a substitute. Though it conveys a correct idea of the general purpose of this circuit, the name has become associated also with at least two other devices used in closely associated electronic fields. (For example, in the terminology of well-known manufacturers, "peaker" denotes a particular stage in a video amplifier designed to emphasize high-frequency components.) The reader is advised to study the RC circuit and ignore the nomenclature. This is the intended significance of the quotation marks enclosing the titles of Secs. 21 and 22.

The assumption that the pulse height is independent of RC naturally fails when the time constant becomes small enough to bring out any departure from accurate "squareness" in the original rectangular alternating voltage. If pulses with a time constant of 1 microsec or less are to be produced, the rise and fall of the rectangular wave evidently must retain their steep and angular character when viewed with an oscillograph capable of resolving such small intervals of time.

In Fig. 21.1 a symmetrical rectangular wave is used arbitrarily as an illustration, and as a result the positive and negative pulses are uniformly spaced. This is not a necessary condition, as the periods A and B , Fig. 21.1, could have been made quite unequal.

The output pulse depends in spacing, polarity, and magnitude upon the corresponding jump in potential of the input wave. It depends in duration upon the RC value of the circuit.

As indicated in Fig. 21.2, unequal positive and negative pulses can be produced by a slanting input wave. Pulses of either polarity may be suppressed by a clipper without difficulty.

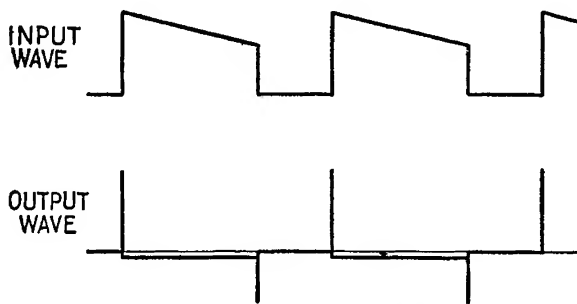


FIG. 21.2.—Pulses of unequal spacing and height.

22. “Integrators.”—Here, again, the established terminology is suggestive but is not to be taken too literally. In a sense the “integrator” is the reverse of the differentiator, the R and C elements exchanging roles and forming an elementary low-pass filter. The name also suggests that emphasis is now to be placed on the cumulative storage of charge in the capacitor C , rather than on the instantaneous response of the current through R . However, unless the time constant RC is extremely large compared with the period of the square wave, the output voltage is not the mathematical integral of the input voltage. The true integral would be a triangular wave.

As with the differentiator, or peaker, the theory of the integrator is completely covered in Chap. VI. In fact, the two circuits are identical except for the position of the output leads. As indicated in Fig. 22.1, the output wave is made up of exponential arcs, approaching a rectangular form as the time constant RC is progressively decreased. However, contrary to the conventional practice in using a differentiator circuit, applications involving the integrator usually require a time constant at least comparable in magnitude with the interval A or B . Ordinarily, it is the relatively sluggish response of the integrator that is needed and is brought out by the detailed design.

For example, the integrator is frequently called upon to respond to a pulse signal of long duration, purposely ignoring short pulses

and other high-frequency components that may precede the proper operating signal. An integrator and a differentiator, fed in parallel by mixed pulses, systematically route the long pulses to one destination and the short pulses to another. In other applications the integrator serves as a delay mechanism, holding back a signal or creating a definitely measured pause while awaiting the beginning or the completion of some other operation. Integrators and clippers often appear in combination, no response whatever being obtained from the clipper until the output wave from the integrator eventually climbs above the threshold of the clipper. The delay may be

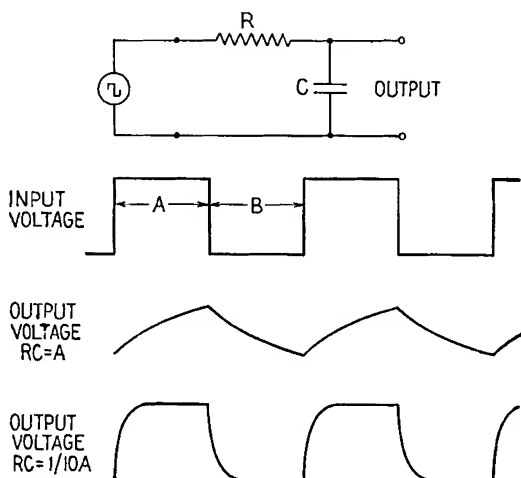


FIG. 22.1.—“Integrator” with various time constants.

adjusted with high precision by altering the time constant RC or by altering the threshold voltage.

23. Saturable-core Reactors.¹—Networks involving saturable-core reactors may be used in place of differentiators for producing short, sharp periodic pulses. The network is somewhat more elaborate, but the required input is an ordinary sine wave rather than a rectangular wave. The resultant pulse is symmetrical and, in uniformity of frequency spectrum, is better than the nonsymmetrical exponential pulse obtained from the differentiator. Special coil construction is required, and the pulse duration cannot be adjusted so easily.

Students are often able to grasp the action of the device readily

¹ E. PETERSON, J. M. MANLEY, and L. R. WRATHALL, Magnetic Generation of a Group of Harmonics, *Bell System Tech. J.* **16**, 437, 1937.

by first considering a comparable problem in mechanics. Assume that a torsion pendulum has been built, Fig. 23.1, similar to the balance wheel of a watch or alarm clock, though it may have been constructed on a much larger scale. A large flywheel rotates slowly on a shaft to which it is rigidly attached. As it rotates, it winds up a strong spiral spring. Rotation, back and forth at the natural frequency determined by the heavy flywheel and the spring, is maintained by an appropriate periodic torque.

Building up by resonance, the inertia and elastic forces involved may become very large. The moving shaft terminates in a flange plate, or coupling plate, moving with the shaft and rigidly keyed to it. Facing directly against this moving plate, but not attached to it, is a similar flange forming the other part of the coupling. This is

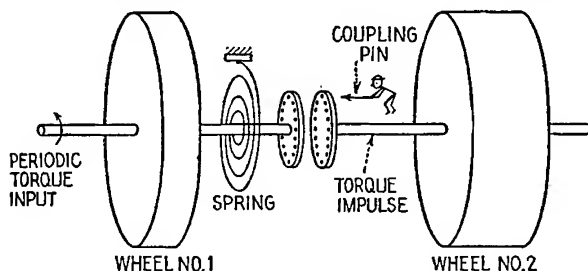


FIG. 23.1.—Mechanical analogue of saturable-reactor circuit.

keyed to a second shaft, for the moment stationary, mounted in bearings that line it up with shaft 1. The second shaft carries a still heavier flywheel, rigidly attached.

At the flange plates stands an ancient "Maxwell's demon," long known to physicists and recently brought to popular attention in new guise as a "gremlin." Watching his opportunity eagerly as the moving flange slows down toward a stop, our little friend deftly slips a small coupling pin through opposing holes in the flange plates just as the relative motion ceases. At this point the spring is fully wound and is exerting its maximum torque. Normally, this torque would be entirely available for accelerating wheel 1. For the moment, however, the same torque must be divided between wheel 1 and the new wheel just added to the system. If this condition persisted, the system would be thrown completely off resonance. However, while attempting to accelerate wheel 2 the small coupling pin promptly shears off, resulting only in the transfer of a momentary torque impulse to shaft 2. Wheel 1 then continues its normal course, and the operation is repeated every half cycle.

The momentary tampering by the gremlin has little net effect on the damping or the period of the moving system, since the actual energy transfer may be made very small. However, by making wheel 2 rather large, the successive positive and negative torque impulses may be made to approach the full torque exerted by the powerfully wound spring at the peaks of its winding cycle. Translated into electrical terms, Fig. 23.2 results.

Capacitor C_1 and coil L_1 form an ordinary series-resonant system tuned approximately to the frequency of the sinusoidal source-voltage e . During most of the cycle the effect of inductor L_2 is negligible. The winding of L_2 contains a considerable number of turns upon a permalloy core designed to saturate at low field intensity. The resonant alternating current has an amplitude far above the saturation level, and hence the flux in general does not alter materially as the sinusoidal current rises or falls through

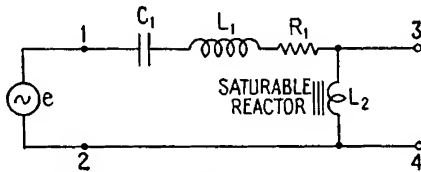


FIG. 23.2.—Saturable-reactor circuit.

a small portion of a cycle. In other words, the self-inductance L_2 is constant and relatively small at the current levels prevailing most of the time. Hence, the small inductance L_2 is not much larger than the inductance possessed by the windings alone, with the permalloy core removed.

As the cycle progresses, however, the alternating current must eventually pass through zero. Remaining unchanged until the current has nearly ceased to flow, the flux in L_2 must first drop almost instantly from its saturation value to zero and must then build up with equal rapidity to a similar saturation level in the reverse direction. This causes a violent pulse of voltage to appear across the coil L_2 every time the resonant current passes through zero. In effect, the coil has suddenly assumed a different value of self-inductance L_2' which may be considerably larger even than that of the fixed inductor L_1 . This high inductance is retained for a brief interval only, during which the flux jumps from one saturation condition to the opposite saturation condition.

The momentary pulse voltage appearing across L_2 is made

possible by the existence of a very high sinusoidal voltage built up across the capacitor C by series resonance. Normally, this is opposed by an equal and opposite sinusoidal voltage at the terminals of the inductor L_1 , which constitutes the only appreciable inductance in the series-resonant circuit. Sudden insertion of the high inductance L_2 forces a temporary *redistribution* of the total inductance voltage, the coil L_2 suddenly demanding the lion's share, instead of the customary mouse's ration. This momentary redistribution takes place without any important alteration in the *total* inductive voltage or the sinusoidal capacitive voltage, which continues to balance it. A slight jog necessarily appears in the current wave as its slope changes for an instant. The sinusoidal

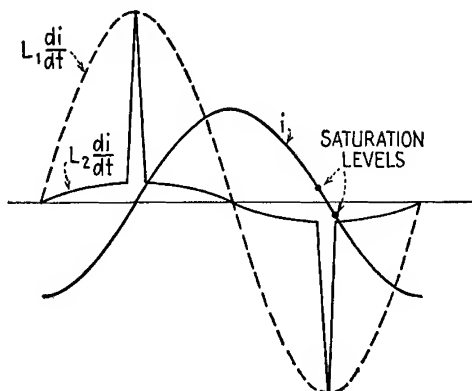


FIG. 23.3.—Voltage pulse.

cycle as a whole is little affected by the sudden insertion and immediate withdrawal of additional electrical inertia. (The mechanical analogue is not offered as an exact counterpart of the electric circuit, though the essential similarity is evident.)

Figure 23.3 represents the waveform obtainable at the output terminals 3-4. As usual, the load circuit should be a high-impedance device drawing little or no current at the frequencies concerned. The sharp pulses may be detached from the sine wave by any appropriate clipper, and the pulses of either polarity may be suppressed if desired.

The circuit finds obvious applications in the production of synchronizing signals, the modulation of pulse transmitters for echo experiments, the production of electronic time scales upon all sorts of oscillograph patterns. It has been applied also commercially to the generation of harmonics. For this purpose, such sharp

symmetrical pulses have the advantage of notable uniformity of harmonic content. For example, using a fundamental frequency of 4 kc, the sixteenth to the twenty-seventh harmonics have been selected by appropriate filters and used as the equally spaced carriers in a multichannel system of carrier-current telephony. At their source, the 12 harmonics in this series differed less than 0.2 db. In this application the odd harmonics were obtained directly from the output terminals 3-4. The even harmonics were obtained through a copper-oxide rectifier bridge. Filtering was facilitated by this separation of odd-harmonic and even-harmonic sources.

24. Keyers, or Electronic Switches.—In elaborate control networks, it is frequently necessary to insert a voltage source, or to remove one, or to substitute one source for another, the changes and substitutions being made periodically in accordance with a definite plan. The changes must be made very swiftly, and the timing must be correct, often within a fraction of a microsecond. Usually the switching device must be synchronized with each voltage source, making the necessary alteration at a carefully predetermined point in each voltage or current cycle. A familiar example occurs when two or more patterns are to be observed, apparently simultaneously, upon a single cathode-ray oscillograph screen. The control leads of the oscillograph are transferred rapidly from one test position to another, while a simultaneous alteration of bias changes the vertical location upon the screen. The cycle of observation is repeated so swiftly that each pattern appears to be present continuously.

In other cases, one particular portion of an observed oscillograph pattern is to be lifted bodily from the pattern as a whole and spread out for detailed examination upon an expanded time scale. A high-speed switching mechanism selects the desired portion of each successive wave. For example, one may wish to examine the double-refraction splitting of one echo selected from an ionospheric reflection pattern. In television practice, the composite synchronizing signal represents the combined output of a number of generators, each having control of the output channel for an allotted number of microseconds during each sixtieth of a second. In certain types of prewar directive devices, built for Coast Guard use, a single receiving set produced two different records upon a common oscillograph screen, the receiver input being shifted rapidly back and forth between two directional antennas. Exact balance of these "split-beam" patterns produced a sharp indication of direction. In ionospheric practice, elliptical polarizations long have been exam-

ined by switching antenna connections smoothly, thus effectively creating a "virtual" antenna that rotates at high speed. Certain modern direction finders utilize a similar technique.

Some of these results conceivably could be secured by mechanical switches, commutators, or relays. Others are ruled out from the start by the microsecond precision required. In all cases, there is a natural trend toward purely electronic operation, with no old-fashioned mechanical movements whatever. As with the clipper, it is important to understand clearly the function of the device rather than to concentrate upon any one specific instrument for performing the operation. The basic requirements are so simple that they may be satisfied in various ways.

To open and close an electronic switch one might naturally consider the possibility of suddenly interrupting and later reestab-

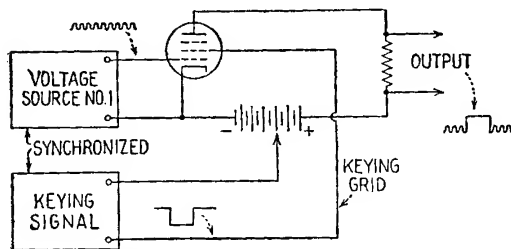


FIG. 24.1.—Keying a signal "out."

lishing the electron stream in a normal amplifier tube. Evidently this could be done by suddenly applying a blocking voltage to the control grid, to the plate, or to any intermediate grid of a multi-electrode tube. In practice, the "keying tube" is usually a tetrode or pentode, and the blocking voltage often is applied to the screen grid, Fig. 24.1.

Voltage source 1 provides a continuous signal of any desired waveform, present at all times upon the control grid of the keying tube. If the screen bias is normal, the "switch" is normally closed, the input signal being amplified in the usual manner. In this case the "keying signal" will be a *negative* rectangular pulse, which suddenly drops the screen potential below the cutoff point, keeping the tube inoperative for the duration of the pulse, as indicated in Fig. 24.1. Under such circumstances, voltage source 1 is said to be "keyed out," leaving a blank space, of the desired length, in the output wave.

Simultaneously, a second keying tube might "key in" a signal

from voltage source 2. Here, a signal of the correct waveform is always present on the control grid, as if knocking for admission at a closed door or gate. The gate is finally jerked open by the keying signal, a positive rectangular pulse that suddenly lifts the screen potential to an operative condition. Having been open for the allotted period the gate then is suddenly closed. The output exists only for the duration of the rectangular pulse. Evidently the same pulse generator can provide both keying signals, applying opposite polarity to the two keying tubes.

Although the keying signals usually encountered are rectangular pulses, occasional applications require a gradual transition from the inoperative to the operative condition, and vice versa. Hence, sine waves, triangular waves, saw-tooth waves, clipped sine waves, and various other pulse forms may be used in special cases. In

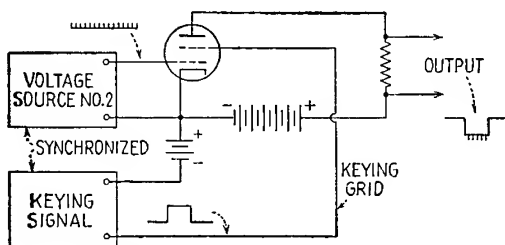


FIG. 24.2.—Keying a signal "in."

effect, these pulses vary the amplification of the keying tube at a predetermined rate. In rather exceptional cases, a remote-cutoff tube may be used in order to accentuate such gradual transitional effects. More commonly, the transition must occur instantly, steepness and precise timing of the keying signal being of major importance. Voltage sources and keying signals usually have a common origin in some master oscillator from which all other signals are derived or controlled. Thus the required synchronization is obtainable, by means to be discussed later.

25. "Mixer" Tubes, or Adding Tubes.—Again the quotation marks in the title indicate that the established terminology should not be presented without a word or two of caution. In discussing superheterodyne receivers, the term "mixer" ordinarily refers to a deliberately nonlinear or modulating device that accepts approximately sinusoidal input voltages of two or more frequencies and yields a complex output wave that includes sum and difference frequencies. However, in the description of timing circuits, the

term usually refers to a linear device that accepts two or more input voltages of completely arbitrary waveform, yielding an output that should be a simple algebraic summation of all such input waves. The function of such mixers might be described less ambiguously by the designation "adding tubes," though this terminology has not received general recognition.

In contrast to arrays of *keying tubes*, which combine several input signals in an orderly *time sequence*, these adding tubes are to combine such input signals as if they were plotted continuously on the same strip of graph paper, the output of the group of mixers being equal at all times to the *instantaneous sum* of these individual ordinates. As in the study of keying tubes and clippers, however, it is

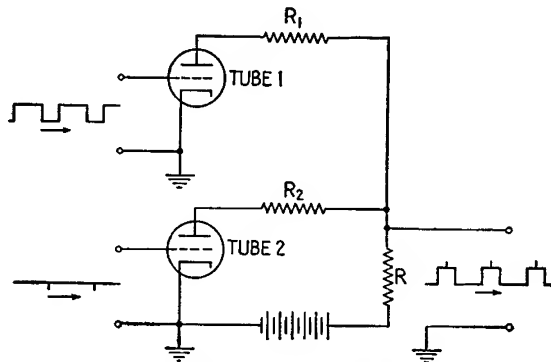


FIG. 25.1.—A simple form of "adding-tube" circuit.

more profitable to concentrate upon an understanding of the prescribed function of the device than to concentrate upon any individual circuit designated as a mixer.

For example, Fig. 25.1 shows a pair of identical triodes, the grids being driven separately by two different component voltages that are to be combined by the mixer. The plate currents of the two tubes are supplied in parallel from a common d-c source through a common coupling resistor R . Hence, if it is assumed that each triode is operated properly as a linear amplifier, a variation in either grid potential produces a proportionate output voltage across R . Simultaneous variations occurring at each grid produce an output voltage linearly proportional to their sum. The auxiliary resistors R_1 and R_2 are a practical convenience, though not a fundamental part of the triode circuit. If these resistors were made zero, the output circuit, seen from either tube, would have an inconveniently

low resistance, enhancing the difficulty of obtaining adequate linearity. This becomes evident when it is noted that the equivalent load circuit, driven by tube 1, would include the variational plate resistance of tube 2, shunted by the resistor R . Hence the effective load upon each triode would be represented by a resistance, necessarily smaller than its own internal effective resistance. This undesirable design condition is easily removed by the introduction of the additional resistors R_1 and R_2 into the network. Conventional amplifier design governs all other details. Naturally, an inversion of waveform takes place as in other simple resistance-coupled amplifiers. Obviously, twin triodes could be used advantageously. Pentodes or tetrodes are often employed in an appropriate amplifier circuit. In using pentodes the auxiliary resistors R_1 and R_2 are not required; because of the high internal resistance of such tubes, the common resistor R is a relatively small part of the total resistance of the circuit. Adequate linearity is obtained easily.

26. Simple Applications and Combinations.—Before examining more elaborate components such as trigger circuits and multi-vibrators, frequently incorporated in a complete timing circuit, consider a few common “one-finger” exercises, chosen in order to illustrate methods of assembling the familiar single-tube components already discussed. The circuit problems quoted below are well known in the art. The solutions presented are by no means original or unique. The orderly step-by-step method of procedure having once been grasped, alternative solutions may be devised readily, some of which may offer special advantages in respect to economy of tubes, precision of timing, or ease and stability of adjustment. The particular solutions quoted are selected to illustrate a variety of functions of numerous simple components.

Problem 1.—Having available a stable sinusoidal timing voltage of frequency 15,750 cps, produce an uninterrupted series of rectangular pulses, each of 10 microsec duration, repeated regularly at the rate of 15,750 pulses per second. Such pulses might be designated as “blanking pulses” and used in turning off the beam of a cathode-ray tube, in order to eliminate a horizontal retrace line.

Solution (for practice, sketch the waveforms and a block diagram, step by step, as the description proceeds):

a. By an amplifier or attenuator, adjust the amplitude of the timing wave to a convenient standard value, say 50 volts.

b. By means of a clipper, cut off and retain the upper 12 per cent of the positive half cycle, discarding the remainder of the timing wave, cutting at the +44-volt level, for example.

c. Amplify the resultant 6-volt curved-top pulses, using a net gain of 20, so that the pulses are stretched vertically to an amplitude of 120 volts, without alteration of their duration. In consequence, the pulse voltage rises from the zero level to the 2-volt level in approximately 0.04 microsec.

d. Clip the amplified pulses at the +2-volt level, discarding the curved top and retaining a nearly rectangular pulse of 9.9 microsec duration, measured along the flat top. The base of the pulse has a duration of 10 microsec. The amplitude may be adjusted at will by further amplification or attenuation.

Problem 2.—Having available the 15,750-cycle rectangular pulse, superimpose on each pulse a pulse of similar amplitude but of much shorter duration, each short pulse being placed 1 microsec after the start of the corresponding rectangular blanking pulse. The short pulses may be designated as “synchronizing pulses” and will be used ultimately in tripping the discharge circuit of a saw-tooth sweep circuit such as the simple relaxation oscillator. En route to this destination, which may be in a television receiver far away, the synchronizing pulses may ride alone on the top of the blanking pulses, thus avoiding the necessity for an extra wire or radio channel. Near their mutual destination the two pulses eventually may be sheared apart by a clipper and thence routed separately to the saw-tooth generator and to the beam-control electrode of a cathode-ray tube.

Solution (continue the sketches of waveforms and block diagram):

a. Having passed the rectangular blanking pulse through a buffer amplifier in order to prevent excessive loading of the circuits discussed in Prob. 1, apply the resultant rectangular pulse to an integrator circuit. Across the capacitor of the integrator the resultant output pulse is made up of exponential arcs, as indicated in Fig. 22.1. In consequence, the leading edge of the pulse no longer rises so steeply.

b. Clip the exponential pulse at an adjustable positive level, adjusted so that the clipping threshold is not reached until the exponential rise has continued for 1 microsec. Discard that portion of the exponential pulse below the threshold level, and retain the portion above. The remaining pulse now begins after the desired 1-microsec time delay, the beginning of the original rectangular pulse being taken as a convenient reference time. However, the delayed pulse does not have yet the desired steepness or short duration. Hence, “pulse sharpening” is next in order and will require several steps, as follows:

c. Amplify the remaining portion of the exponential pulse so that it is stretched vertically to an amplitude of 100 volts or more. In consequence, the rise in voltage, though exponential, brings the voltage from zero to the +2-volt level in a small fraction of a microsecond and may be regarded as substantially vertical over this limited range.

d. Clip this amplified pulse at any convenient low threshold such as +2 volts, thus producing a flat-topped pulse, of low amplitude, having a very steep leading edge, this edge being delayed 1 microsec with respect to the reference time. The pulse does not meet yet the specifications, as it is too long. Since it is to be shortened in the next operation, the exact present position and steepness of the trailing edge will be of no ultimate consequence.

e. Apply the flat-topped pulse to a differentiator, or peaker, circuit, the time constant, RC sec, being small enough to produce a short, sharp positive

pulse that begins 1 microsec after the reference time and thus meets the specifications for the synchronizing pulse. In addition, however, the trailing edge of the applied flat-topped pulse necessarily produces a negative output pulse, having an area equal to the area of the desired positive pulse, though usually of somewhat different amplitude and shape. This undesired negative pulse is to be removed in operation *f*.

f. Pass the sharp pulses through a clipper circuit, slicing the wave at the zero axis. The negative pulses are to be discarded, and the short, sharp positive pulses retained. If necessary, the amplitude of the positive pulses may be adjusted by an attenuator or linear amplifier.

g. Finally, the synchronizing pulses, completed in operation *f*, are to be fitted in their assigned positions on top of the rectangular blanking pulses, which are still available at the buffer amplifier mentioned in operation *a*. Hence the voltages, picked up from these two positions in the circuit chain, may be combined in a mixer, or adding, circuit, thus completing this series of operations.

Problem 3.—Having available the 15,750-cycle rectangular pulse, produce a series of rectangular pulses of identical frequency but of 12 microsec duration, each of the new pulses beginning at the same time as the corresponding 10-microsec pulse. In other words, a “pulse stretcher” is desirable that can extend the duration of rectangular pulses without altering their steepness or the precise timing of their leading edges. In conventional television practice, for example, the blanking pulses sent to the receivers are slightly longer than the blanking pulses used at the transmitter, just as the negatives used in ordinary photographic printing are slightly larger than the printing mask and the resultant contact print. In each case, this slight margin, or tolerance, simplifies adjustment and ensures a picture with sharp and clean edges.

Solution (continue the sketches of waveforms and block diagram):

a. As a first step, a delayed pulse is desired, the leading and the trailing edge each being moved backward in time. This can be accomplished by an

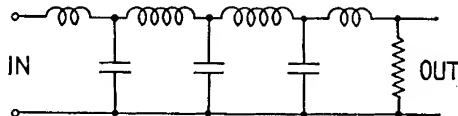


FIG. 26.1.—Delay circuit.

integrator and subsequent clippers, precisely as in Prob. 2. To avoid mere repetition, however, an alternative method yielding similar results may be employed. On this occasion, therefore, the original rectangular pulses are applied to the input terminals of a “delay circuit,” or “lag line,” Fig. 26.1. This is a ladder network consisting of series inductors and shunt capacitors, or it may consist of series resistors and shunt capacitors. Usually two, three, or four sections suffice for delaying a rectangular pulse. Although each section of the artificial line does introduce a small time delay as desired, the device also functions as a low-pass filter, which practically eliminates all harmonics of the input signal above the cutoff frequency. Such filter action might, therefore, spoil the waveform of the rectangular pulses. In practice, however, a satisfactory engineering compromise is feasible. The pulse repetition frequency

seldom exceeds 20,000 cycles. Even in this extreme case, harmonics as high as the fifteenth may be preserved in the delayed pulse if the cutoff frequency be set at 300,000 cycles or more. Using coils and capacitors of the order of 1,000 μ h and 1,000 μ f, this requirement may be met while introducing a time lag of approximately 1 microsec per section. Use of an *RC* ladder network would also be satisfactory from the time-lag standpoint. Greater attenuation would result, but this may be offset by amplification.

b. By means of a linear mixer circuit, add the delayed rectangular pulses to the original rectangular pulses, the two input voltages being adjusted to equal amplitudes. The resultant pulse now has the prescribed length and correct timing, but it does not have a continuous, flat top. Instead, it has a stepwise waveform, similar to the profile of a tall building with equal architectural setbacks at front and rear. In the period of overlap, while both input pulses are controlling the mixer, the output pulses necessarily have a double amplitude.

c. By means of a clipper, trim off the unnecessary superstructure, thus leaving the desired extended rectangular pulse, which begins simultaneously with the original rectangular pulse but does not end until the delayed pulse terminates. To provide a convenient engineering tolerance, the threshold is to be adjusted so that the clipper will shave off a small portion of the top of the single-amplitude part of the stepwise pulse and therefore will completely and clearly remove the double-amplitude projection. Any slight loss of amplitude incurred in providing this margin may be restored by further amplification or may be allowed for in setting the gain controls previously encountered.

The common exercises just quoted are sufficient to indicate the step-by-step arrangement that characterizes a typical timing circuit. Though the individual steps are simple and familiar, they may be very numerous, and in the aggregate they necessarily will require a multiplicity of tubes. To a considerable extent the number of tubes may be held within bounds by frequent recourse to twin triodes and twin diodes. At the relatively low frequencies concerned, shielding problems are not difficult. As practical applications of such elaborate circuits become increasingly numerous, it is possible that manufacturers may eventually include still larger numbers of electronic units within a single glass envelope, with resulting advantages in compactness and efficient use of power.

27. Trigger Circuits.—Trigger circuits are electron-tube networks that possess two or more stable operating conditions, to which the network will automatically return after experiencing minor disturbances of internal or external origin. Thus ordinary noise voltages, of moderate amplitude, have no resultant effect upon the network. However, a *sufficiently strong* transient voltage, deliberately introduced from an external source, may momentarily drive the circuit so far as to lift it over the boundary of a

particular state of equilibrium, whereupon all the currents and voltages slide rapidly toward a *new* set of stable-equilibrium values. In many cases the circuit is permitted initially to select at random any one of its stable conditions, the chance decision being made immediately after the circuit is completed with all the power supplies in place. Thereafter, the successive transitions from one stable condition to the next take place in a regular predetermined sequence, each operation altering the pattern of steady potentials and currents in such a way as to determine the location and nature of the next event. However, the *timing* of the next transition depends entirely upon the external control, any consistent and stable set of currents and voltages being held for an unlimited time while awaiting the next triggering impulse from the external source. In consequence, operation in proper sequence is *not* dependent upon availability of evenly spaced driving pulses. In fact, in some applications great irregularity in the arrival of these control impulses is expected. Large fluctuations in *amplitude* of control signals may also be tolerated without disadvantage so long as the amplitude never drops below a definite minimum level. Such circuits may be compared to a mechanical ratchet and pawl mechanism and may be used for similar purposes. Their major advantage in comparison with a mechanical-ratchet relay lies in the high speeds obtainable without error. Two driving pulses, spaced 10 microsec apart, may be accepted as separate events and may be made to produce the intended full sequence of operation. This is about 1,000 times faster than the standard speeds of comparable electro-mechanical devices. Still higher resolving power can be obtained by careful design but ordinarily is not needed.

These general principles may be illustrated by considering in some detail the operation of the Eccles-Jordan¹ trigger circuit, sometimes graphically referred to as a "flip-flop" circuit. Several other basic circuits have been invented and employed, but the Eccles-Jordan circuit is the oldest basic design and is widely used at present. Like the Wheatstone bridge, it may be regarded as a prototype and as a very rich source of ideas, numerous modified and specialized circuits having been derived from it. In fact, the same diamond-shaped pattern made familiar by the Wheatstone bridge appears also in this basic trigger circuit, Fig. 27.1, and in many of its offspring. In this case, two similar vacuum tubes

¹ W. H. ECCLES and F. W. JORDAN, A Trigger Relay Utilizing Three-electrode Thermionic Vacuum Tubes, *Radio Rev.*, **1**, 143, 1919.

form adjacent sides of the rhombic pattern, their external plate-load resistors completing the four-sided net. The Wheatstone bridge analogy must not be pursued too far, however, as the cross connection of the grids deliberately prevents establishment of a stable balanced condition of equal or nearly equal currents flowing in the symmetrical branches of the net. In order to demonstrate this intentional instability a complete cycle of operation may be traced as follows:

Assume, as usual, that the plate power source has been disconnected from the circuit, that sufficient time has elapsed so that all transients have died away, and that the plate power supply is then suddenly restored. In view of the obvious symmetry of the circuit, with completely similar tubes and resistors at all correspond-

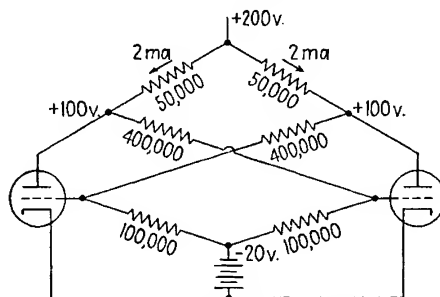


FIG. 27.1.—Trigger circuit: unstable equal-current condition.

ing positions on the left and right sides of Fig. 27.1, the trial assumption may be made that equal currents flow through each tube and its associated circuits, determining a common value, say $+100$ volts, for the plate-to-cathode potential of each tube. In fact, this would be a possible and a natural equilibrium condition at the instant when the power supply is restored. This balance can endure but a moment, however, as it is violently unstable. Minor random fluctuations are continually occurring because of "hiss noise." At frequent and irregular intervals, additional "noise voltages" are generated by positive ions, leakage fluctuations, etc. Immediately after the power supply is connected, some chance fluctuation will cause a minor unbalanced variation in the current of the left tube or the right tube. To be specific, let us assume that the current of the left-hand tube happens to increase momentarily. Because of the cross connection of the grid circuits, this slight initial variation sets off a rapidly cumulative process, which immediately exaggerates the first minor change. The slight initial increase of

current flowing through the left-hand plate-load resistor produces a minor drop in the corresponding plate voltage, bringing it, say, 0.1 volt below the 100 volts previously mentioned.

Note that the 400,000-ohm and 100,000-ohm resistors form a pair of potential dividers, Fig. 27.2. Consider these dividers, shown separately in the figure for special emphasis. By means of a battery or any equivalent low-resistance d-c source, the lower terminal of the potential divider is held securely at a fixed potential, say -20 volts. The upper terminal is attached to the plate of the left-hand triode.

A slight reduction in the grid current flowing through the 400,000-ohm resistor being neglected, a momentary reduction of 0.1

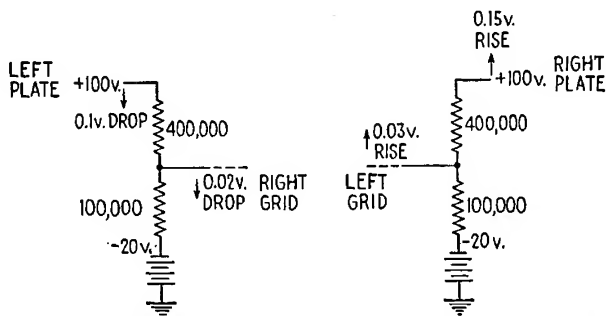


FIG. 27.2.—Potential-divider action of the coupling circuits of Fig. 27.1.

volt in the potential of the left-hand plate produces a proportionate 0.02-volt reduction in the potential of the right-hand grid. The momentary lowering of this grid potential raises the right-hand plate voltage, say by 0.15 volt. In turn, approximately one-fifth of this rise is passed back to the left-hand grid. If the resultant 0.03-volt rise should happen to be just sufficient to sustain the original increase in the plate current of the left-hand tube, then the system would be in neutral equilibrium and the initial accidental change would be self-perpetuating. In practice, however, the design of the circuit is such that the reaction provides an "overpayment." The initial current increase is not merely sustained but is augmented. The action being cumulative, the system immediately "runs away" in a violent and very rapid transient, which sends the left-hand plate current rocketing upward toward a maximum, while the right-hand plate current drops quickly toward zero. These rapid changes stop wherever the reaction, described above, ceases to produce the overpayment. As in the case of the blocking oscil-

lator, there is a ceiling, imposed by Ohm's law, above which the plate current cannot increase. In Fig. 27.1, for example, the 200-volt battery could not deliver more than 4 ma through the 50,000-ohm load resistor, even with the tube short-circuited. As the current approaches this value, the chain of events described yields progressively smaller increments in the potential of the left-hand grid. Eventually, the final increment of grid voltage is barely able to sustain the corresponding increment in plate current, with no margin for further increase.

Let it be assumed arbitrarily that this condition is fulfilled when the current drawn by the left-hand tube has increased to about 3 ma, Fig. 27.3. At this point, all changes of current and voltage cease suddenly, the system having reached a new equilibrium.

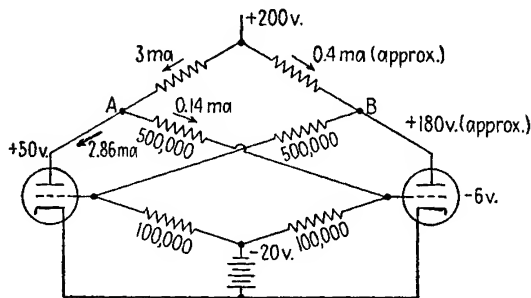


FIG. 27.3.—Trigger circuit: stable terminus condition.

Unlike the unstable equilibrium, characterized by momentary equality of current flow, this new equilibrium condition is completely stable with respect to small variations in voltage or current. This important point may be explained as follows:

Let the total current flowing in the left-hand load resistor be 3 ma, equal to the sum of 2.86 ma of plate current and 0.14 ma of current in the potential divider. Allowing for 150 volts drop in the 50,000-ohm external resistor, the plate of the left-hand tube is only 50 volts above ground or cathode potential. The entire difference of potential across the first potential divider is therefore 70 volts, one-fifth of which, 14 volts, appears across the 100,000-ohm lower section of the divider. Consequently, the grid of the right-hand tube is held at a potential of -6 volts, this being 14 volts above the -20 level. This grid voltage is sufficiently negative to ensure that the plate current of the right-hand tube will be very small if not actually zero. In other words,

the right-hand tube is biased near the cutoff level, in some cases below the cutoff level. Consequently, the right-hand plate potential has risen nearly to the level of the plate-supply voltage. The drop in voltage may be due to the current drain of the potential divider alone and may therefore be as low as 20 volts. The potential of the right-hand plate is approximately 180 volts above ground. Hence the voltage across the second potential divider extends from $+180$ to -20 . If the grid current of the left-hand tube could be neglected, this grid would be held at 40 volts above the -20 level, a net potential of $+20$ volts. Though this highly positive value is reduced nearly to zero by the actual grid current superimposed upon the network, the grid potential remains slightly positive, being high enough to permit the 2.86 ma to flow through the tube with an internal voltage drop as low as 50 volts. Hence the entire system of highly unequal currents and unequal voltages is now self-consistent and self-perpetuating, there being no unbalanced voltages tending to produce or assist any further change.

Moderate fluctuations of voltage or current now can be tolerated and will provoke no significant reaction from the circuit. For example, a minor decrease below the assumed 2.86 ma in the left-hand tube merely restores proportionate parts of those same unbalanced voltages which originally brought the circuit to the "terminus" condition at the end of its rapid transient sweep. The unbalance being dependent upon the instantaneous values of the currents and voltages and not upon their time derivatives, the circuit immediately drops back to the same terminus condition. On the other hand, any momentary increase above the assumed 2.86-ma equilibrium is unfruitful, the alteration of grid voltages being insufficient to maintain the momentary increase. Hence the circuit again drops back to normal and is stable with respect to minor fluctuations of arbitrary direction, location, and character.

There is, however, another equally good condition of stable equilibrium, in which the two tubes simply interchange their roles. Starting from the equal-current condition, the unstable "seesaw" circuit could have tipped just as readily in the *opposite* direction and would have done so if the initial accidental variations had happened to produce a momentary *decrease* of the left-hand tube current or a momentary increase of the right-hand tube current. When the plate power supply is first connected, the selection of one of the two possible stable conditions is purely a matter of chance.

The chance decision, having quickly been made the corresponding pattern of steady currents and potentials should be held indefinitely, until finally terminated by removal of power supply or by a sufficient disturbance deliberately introduced from an external source.

At regular or irregular intervals the flip-flop circuit is to be *driven* back and forth from one stable condition to the other, but every reversal is to occur under the complete control of some external device. As soon as the control voltage ceases, the circuit becomes quiescent and remains in the condition determined by the last reversal.

The process of forcing these sudden transitions from one stable condition to another is called "triggering." In the simple Eccles-Jordan circuit, triggering may be accomplished in a number of different ways, but in all cases the triggering voltage temporarily forces a reduction in the great inequality of the two plate currents. If this inequality can be reduced to zero and then reversed in sign (by any marginal amount, however small), the triggering is successful; the circuit automatically proceeds further and arrives at the opposite terminus condition. However, if the triggering voltage is not quite strong enough to drive the inequality to zero, the triggering is ineffective; the circuit flops back into its previous stable condition, just as a ladder, slowly lifted from the ground toward a vertical position but released before the vertical position is quite attained, will drop back to its previous horizontal position.

An elementary but rather impractical method for triggering the circuit may be developed by forcing additional currents through the load resistors, the additional currents being controlled by some external device. For example, an external battery with negative terminal connected to point *B* and positive terminal to point *A*, Fig. 27.3, sends current through the network in such a direction as to reduce the inequality of plate potentials and the consequent inequality of grid potentials and tube currents. Progressively increasing the battery voltage forces the entire system nearer and nearer to the equal-current condition. On finally passing through this unstable equilibrium position, the system suddenly reacts violently as it snaps over into the alternate terminus condition. Triggering has been accomplished by brute-force methods. A complete circuit of this plate-driven type has been described by Stevenson and Getting and employed in cosmic-ray research. The main disadvantage is the necessity for using amplifier tubes in parallel with each tube of every trigger-circuit pair. As a complete

control circuit often employs many trigger-circuit components, the array becomes needlessly cumbersome.

In using triodes in an Eccles-Jordan circuit, provision for triggering often is made by attaching small coupling capacitors (say $25\ \mu\text{f}$) to each grid, Fig. 27.4. Assume that a negative pulse, derived from an external source, causes a sudden and temporary drop in the potential of terminal 1. Because of the resistance in series with the capacitor C_A , any charge on the plate of this capacitor cannot be altered instantly. Hence the drop in the potential of point 1 produces a corresponding drop in potential at $1'$, no alteration in the potential difference *between* the plates of the capacitor being possible in this first important instant. If grid $1'$ happens to be blocked below cutoff, the tube does not respond, the negative

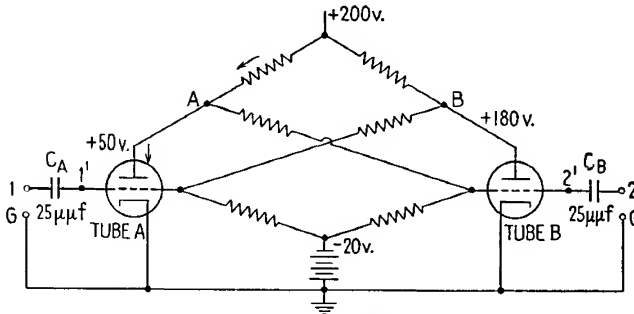


FIG. 27.4.—Separate triggering controls.

pulse being disregarded. However, if tube A has been carrying the heavy current, the sudden reduction of the positive potential of its grid causes a sudden and violent drop in current accompanied by a simultaneous rise in the current of tube B . If the instantaneous current values *meet and pass*, the circuit triggers, flipping over to the opposite terminus condition. If the reduction of current A and the consequent rise of current B are insufficient, the circuit fails to trip, the effect of the triggering pulse being too weak. Note that insufficient pulse strength at $1'$ may be due to insufficient pulse amplitude at the source or to lack of steepness of the pulse wave front. A sloping wave front necessarily permits a portion of the available voltage to be expended in the series element C_A , with a decreased pulse available for use at $1'$. The capacitor C_A might be made larger, were it not for the fact that its twin C_B ties the grid $2'$ to the inactive potential source 2. This is a disadvantage, as an alteration in the charge of C_B must accompany the desired rapid

rise in potential of $2'$. Too large a value of capacitance decreases the sensitivity by limiting the suddenness of the rise of current in the plate resistor connected to point B . The best compromise is obtained by making the triggering pulses as sharp as may be feasible (say a 0.1-microsec wave front) and then reducing the coupling capacitance until the sensitivity is a maximum (say $25\ \mu\text{f}$). In general, decreased steepness of wave front necessitates an increased pulse amplitude with larger coupling capacitors.

Assuming that negative pulses only are received from an external source, the circuit of Fig. 27.4 will respond to the first negative pulse, of sufficient strength, arriving at terminal 1. Having switched the main current over to path B , the circuit then disregards any additional negative pulses imposed at point 1, while

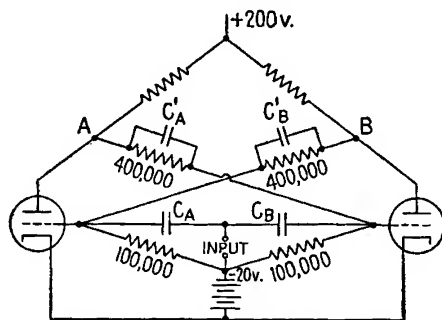


FIG. 27.5.—Single triggering control.

awaiting the occurrence of the next negative pulse applied at point B . When this arrives, the circuit switches back to the condition of Fig. 27.4.

In many applications, the points 2 and 1 are joined together, pulses from a single external channel being applied simultaneously to each coupling capacitor, Fig. 27.5. Addition of "commutating capacitors" C'_A and C'_B prevents the circuit from stalling when equal negative pulses are applied simultaneously to coupling capacitors C_A and C_B . For example, suppose that tube A is conducting. Application of a negative pulse at the common input terminal causes a sharp drop of plate current through tube A . The resultant sudden rise of plate voltage transferred through C'_A to the grid of tube B overpowers the negative pulse, of external origin, that is applied through C_B to the grid of tube B . The first pulse flips the main current over from A to B , the next pulse flips it back, and the operation proceeds at a regular or an irregular rate, dependent

entirely upon the spacing imposed by the pulse source. The size of C_A' and C_B' is not critical. Usually they are twice as large as C_A and C_B .

Positive pulses, interspersed with negative pulses, will cause additional reversals of the current values, if the positive pulses are sufficiently powerful. In this case, it is the blocked grid that responds strongly, the conducting tube being relatively insensitive to a momentary further increase in its positive grid potential. The pulse amplitude required for positive triggering is ordinarily somewhat greater than that required for negative triggering. The difference in sensitivity is not adequate for dependable discrimination of polarity, however. Hence, when operation is to be con-

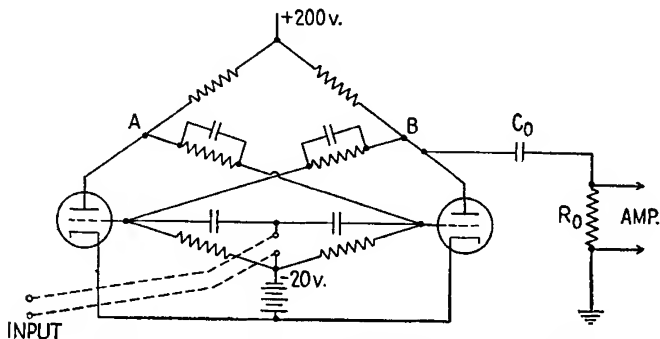


FIG. 27.6.—One method of obtaining an output voltage.

trolled by negative pulses only, it is customary to trim off with a conventional clipper any positive pulses that might otherwise cause erratic operation.

Output voltage may be obtained by attaching a coupling capacitor to any point whose potential varies with time. Sometimes a coupling resistor, say 1,000 ohms, is inserted in the cathode circuit. In this case, a rectangular voltage output is obtained, proportional to the corresponding tube current. A similar wave may be obtained by connecting a large coupling capacitor C_o to point A or to point B, Fig. 27.6, in which case the potential falls sharply as the corresponding tube current rises. If the time constant $C_o R_o$ be large in comparison with the longest expected interval between pulses, the output voltage will be approximately rectangular. Preferably the trigger circuit should be followed by a buffer amplifier.

To go to the opposite extreme in design, the time constant $C_o R_o$ may be made very short in comparison with the smallest expected

pulse interval. The output circuit thus becomes a differentiator, or peaker, yielding a sharp positive pulse when the current at *B* falls, followed by a sharp negative pulse when the current rises. Thus, for every pair of negative pulses arriving at the input terminals, two pulses appear at the output position, but only *one* of these output pulses is negative. Hence, regardless of possible regularity or irregularity in the spacing of the input pulses, every trigger-circuit pair may be employed to cause a reduction *by a factor of two* in the number of negative pulses routed through the device. Several applications are based upon the operation of such trigger-circuit pairs in cascade, constituting a linear chain in which each trigger-circuit stage is driven by the preceding stage. The buffer amplifiers between stages also function as clippers, disposing of the unneeded positive pulses.

Used as a counter of cosmic rays or other irregular events, the circuit greatly reduces errors due to the occasional occurrence of nearly simultaneous pulses. Being able to respond approximately 1,000 times as fast as a good mechanical relay, the circuit may be made to trigger twice when driven by two pulses spaced 10 micro-sec apart. If the number of pulses is scaled down by a factor of 2 at each stage in the cascade system, the cosmic-ray counts may be delivered finally in batches of 16, or 32, or 64, to a relatively crude mechanical recorder, the statistical probability of overhurryng the mechanical ratchet being reduced to any desired extent.

Note that such a chain, made up of 5 trigger-circuit pairs, may be regarded as a *single* electronic network having 32 (that is, 2^5) different stable combinations of current and voltage, each different combination extending over the entire 10 trigger tubes of the system and their auxiliary buffer amplifiers. In fact, these 32 different combinations may be exhibited readily by incandescent lamps, neon bulbs, meters, or any other devices that will indicate at a glance whether the left tube or the right tube of any pair is carrying the main current. Hence, whenever the system is quiescent between counts, the pattern of lights indicates the exact number of counts received since the last batch of 32 was delivered to the mechanical recorder. Each of these 32 patterns also predetermines the location and the extent of the changes that *will* take place whenever the next input pulse arrives. Hence, each of the 32 patterns is occupied in a regular sequence even though the sudden transitions may occur at a wholly irregular rate completely determined by the timing of the input pulses.

An important special case arises when the input pulses are *regularly* spaced, being derived, for example, from a local oscillator. The counting chain now functions as a reliable frequency reducer. Unlike other frequency reducers, it has no natural, or preferred, frequency of its own and is therefore able to follow variations in the frequency of the control oscillator without error even though these variations may occur very rapidly. The chain may be compared to a reduction gear train in which the input shaft speed is arbitrary and may be allowed to vary quickly from a full stop to a speed of 100,000 cps. Some other types of frequency-reducing device, such as the multivibrator, are designed for some particular set of input and output frequencies. Very minor deviations of input frequency are tolerated provided that they occur smoothly, but major alterations of shaft speed are not acceptable. The fact that each pair of trigger tubes divides by 2 and only by 2 is an obvious limitation, though this very feature promotes reliability of operation. Simple linear chains of trigger pairs must divide by powers of 2, such as 2, 4, 8, 16, 32, etc.

If a considerable investment in tubes is warranted for a special application, this limitation may be removed. In place of the linear chain we may construct a circular chain, each link in the system consisting of a pair of trigger tubes. The chain forms a closed ring, completely symmetrical, having no beginning or end. The control grid of the left-hand tube of every pair is connected through a small coupling capacitor to a common input terminal. Negative pulses from an external source are transmitted by these capacitors simultaneously to the left-hand tube in *each* trigger pair. However, in all pairs *except one*, the left-hand tube is carrying no current and hence is insensitive. The *single* sensitive pair responds by the usual exchange of currents, thereby desensitizing itself, while at the same time it produces a negative pulse upon the right-hand grid of the next pair in sequence. The consequent reversal of this adjacent pair completes the transient response initiated by arrival of one negative pulse at the common input terminal. While awaiting the arrival of the next control pulse, the entire ring circuit remains in a stable, quiescent condition, with one pair sensitive and all other pairs insensitive. If an indicating lamp is placed in every left-hand plate circuit of a 10-pair ring, the light rotates stepwise around the circle, requiring 10 input pulses for each round trip. Output voltage, derived from cathode, grid, or plate of an arbitrary pair, has the desired tenfold reduction in frequency. The circuit has 10

stable-equilibrium conditions and occupies each in sequence. Any other integer may be chosen, and a cascade connection is entirely feasible. This electronic ratchet may be used also as an electronic commutator with an arbitrary number of evenly spaced brushes. For example, a 3-pair ring may be used as a 3-position electronic switch, throwing 3 separate patterns on an oscillograph screen in rapid succession, each pattern occupying exactly $\frac{1}{3}$ of each complete ring cycle. As in the conventional 2-position electronic switch, the relative sizes and positions of the 3 patterns may be altered at will.

Trigger-circuit systems of the Eccles-Jordan type by no means are limited to triode tubes. Using pentode tubes, for example, the suppressor grids may be resistance-coupled to the opposite plates, as in the triode circuits previously described. Appropriate grid-bias voltages and grid-leak resistances determine a pair of stable terminus conditions, either of which may be occupied. Used in this manner, the suppressors may be referred to as "holding" grids. In contrast, the control grids become "tripping" or "triggering" grids, charged only with the function of *cutting off* the electron stream on arrival of a negative pulse of external origin, thus causing transition from one terminus condition to the other. (Positive pulses could also be employed, the nonconducting tube being sensitive.) This separation of the functions of tripping and holding is a real advantage, increasing the sensitivity. The holding grids can respond readily, being free of attachments to external circuits. The pulse source affects the electron stream directly, controlling a grid used for no other purpose. Other combinations of grids may be utilized, maintaining the same desirable separation of purpose. To the same end, some improvement in the sensitivity of the triode circuit is occasionally sought by the use of small copper-oxide rectifiers¹ in place of the input coupling capacitors C_A and C_B of Fig. 27.5. In this case, it is intended that the external circuit shall trigger the sensitive grid without interfering with the resulting transient. Use of a pentode is in general a better solution.

Special applications, not concerned with frequency reduction, are too numerous to mention. Brief reference to one such application may serve as an illustration of the versatility of the device. For example, one may wish to construct a phasemeter capable of operating within 2-degree accuracy at all frequencies up to 10,000 cycles. The meter shall respond quickly so as to follow rapid

¹ W. B. LEWIS, *Proc. Cambridge Phil. Soc.*, **33**, 549, 1937.

alterations of phase, shall be independent of amplitude within reasonable limits, shall record phase changes graphically when desired. The problem is solved as follows:¹

Given two approximately sinusoidal voltages whose relative phase angle is to be determined. Using amplifiers or attenuators, bring their amplitudes to the same order of magnitude, equality being unnecessary. By successive clipping, previously described in Sec. 17, square these voltage waves; then differentiate the resultant rectangular waves; clip off the positive peaks. Sharp negative peaks now mark the positions on each time axis at which each of the original sine waves passed through zero, headed downward. Couple one of these pulse sources to the left-hand grid of a trigger-circuit pair. Couple the other pulse source to the right-hand grid. If the original voltage waves happen to differ in phase by exactly 180° , the main current will flow for equal intervals of time in each tube. In general, this is not the case, and the operating periods are therefore unequal. As the current waves within the trigger circuit are rectangular, a d-c meter connected in the proper cathode lead reads an average current linearly proportional to the phase angle. Accuracy is satisfactory, except near 0 and 360° . This is not a limitation, however, as an inverter stage readily shifts one of the voltages 180° whenever desired. The same inverter permits immediate self-calibration of the phasemeter, a single voltage source being compared with its own inverse wave, separated by the known 180° . Phase fluctuations may be followed as swiftly as the d-c meter will permit. A recording ammeter may be used when required. The method has been extended² to the comparison of radio-frequency voltages and has been employed successfully in determining angles of arrival of incident radio waves. Two r-f voltages, picked up by spaced antennas, produce audio-frequency voltages by beating with a common local oscillator. The original phase differences are preserved in the beat notes and can then be accurately measured.

Similar special applications include frequency meters and various devices for measuring mechanical speed by means of photocells, magnetic pickups, and the like.

28. Multivibrators.—The ordinary sinusoidal oscillator may be regarded usefully as a self-driven amplifier, often of the Class *C* type. In similar fashion, the multivibrator may be treated as a

¹ J. E. SHEPHERD and F. M. WIENER, unpublished Cruft theses.

² JOSEPH KEARY, unpublished Cruft thesis.

self-tripping trigger circuit, closely related to the basic Eccles-Jordan device. This relationship is emphasized by drawing the multivibrator circuit in the symmetrical pattern of Fig. 28.1. The coupling capacitors C_A' and C_B' have replaced completely the coupling resistors of the Eccles-Jordan circuit; the grid-bias power source is removed; no other modifications are required. The

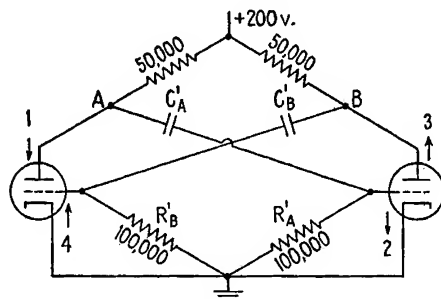


FIG. 28.1.—Basic multivibrator circuit. The numbered arrows indicate the sequence of changes in potential as described in the second paragraph of Sec. 28.

multivibrator may be given quite a different *appearance* by drawing it in the form of a two-stage resistance-coupled amplifier with feedback, Fig. 28.2. Note that these two configurations are in fact identical. This basic multivibrator circuit will be described as a “symmetrical multivibrator.” The previous discussion of trigger circuits, Sec. 27, now will be extended to include this self-triggering function.

In particular, the discussion of the violent instability of the equal-current condition applies also to the multivibrator. On

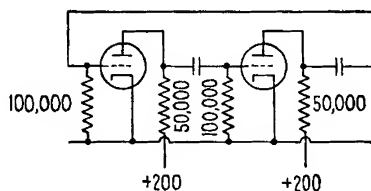


FIG. 28.2.—Basic multivibrator circuit.

suddenly completing the power circuit after a period of inactivity, approximately equal currents might flow for an instant through the two tubes in their completely symmetrical circuits. However, a chance fluctuation may cause a very slight increase in the current of the left-hand tube. The resultant drop in voltage at point A, Fig. 28.1, depresses the potential of the opposite grid, the potential

difference across C_A' remaining fixed for the moment. The decrease in current through the right-hand plate-load resistor causes a rise in potential at B , and this in turn draws the left-hand grid upward. Therefore the initial current increase is augmented. The electronic seesaw mechanism is in full swing, driving the currents in opposite directions from the equal-current condition. Current in the left-hand tube approaches the ceiling imposed by Ohm's law. Current in the right-hand tube immediately drops to zero. Flow of grid current causes a rapid adjustment of the potential difference across C_B' limiting the rise of potential at the left grid. In contrast, the charge upon the plates of C_A' cannot change rapidly. Therefore the right-hand grid is driven far below cutoff, blocking the tube for a considerable time. Thus far the action parallels the corresponding trigger-circuit behavior.

A new feature now appears. In the Eccles-Jordan circuit the terminus condition is a stable, quiescent state, retained indefinitely until interrupted by external force. In the multivibrator this quiescent condition is impossible. The grids are connected to the cathodes by the grid resistors; they are connected to other points in the circuit by capacitors only. Given time enough for the capacitors to charge and discharge, the grid potentials cannot be held steady at any value other than zero. As the left-hand grid voltage drifts downward toward zero, the corresponding current decays gradually from its high initial value. Meanwhile, the right-hand grid gradually rises toward zero from a value far below. In so doing it eventually crosses the "cutoff" threshold. This constitutes the act of pulling the trigger and releasing the next violent phase of the transient. The first weak trickle of electrons reestablishing the current flow through the right-hand tube loads the right-hand plate resistor slightly and lowers the voltage of point B . Acting through coupling capacitor C_B' , this hastens the downward drift of the left-hand grid voltage and plate current. In turn, this speeds the rise of potential at point A . Through coupling capacitor C_A' , this accelerates the upward drift of grid voltage already in progress at the right-hand tube. All these actions being strongly cumulative, all slow drifts of current and voltage are suddenly converted into powerful and almost instantaneous drives. All waveform traces turn through sharp corners and become approximately vertical, Fig. 28.3. The right-hand tube suddenly draws maximum current, while the left-hand tube becomes blocked, with a generous margin of blocking voltage to spare. The relatively

slow drifting process then resumes, and this margin is gradually cut down until it vanishes. At this time, the left-hand grid triggers the entire circuit, and the two tubes again exchange roles suddenly.

Evidently the crucial factor in determining the duration of such a cycle is the rate at which the grid of the blocked tube drifts back toward the cutoff threshold. Control of the next triggering operation is handed back and forth from one tube to the other, control always residing in the tube that is blocked at the moment. As the individual tubes exert separate control over the "halves" of the cycle, it follows that these halves are not necessarily equal.

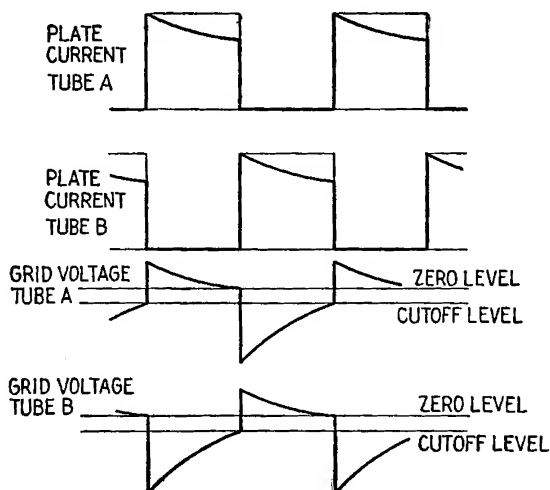


FIG. 28.3.—Multivibrator waveforms in the circuit of Fig. 28.1.

In fact, in many specialized circuits advantages are obtained by deliberately making the two portions of the cycle unequal, sometimes by a large ratio such as 100 to 1. Such circuits may be referred to as "unbalanced multivibrators." Further specialization sometimes leads to a difference of configuration as well as a difference of constants between the left and right side of the multivibrator. Such "asymmetrical multivibrators" follow the general cycle of operations described in spite of their apparent dissimilarity. Hybrid combinations of a trigger circuit and a multivibrator are also of interest. In such a circuit the coupling capacitor C_B' might be removed and replaced by a coupling resistor R_B , the coupling capacitor C_A' remaining in place, Fig. 28.4. In this case, the circuit returns automatically to a single-terminus condition, executing one complete cycle whenever triggered by an outside pulse. Similar

action results when a negative bias, sufficient to maintain cutoff, is introduced in *one* grid circuit of an otherwise conventional multivibrator (circuit derived from Fig. 28.4 by replacing R_B with C_B').

In specialized multivibrators, as in the basic symmetrical circuit, the time constant of each blocked grid is of major importance. Each time constant is a product CR , where C is a coupling capacitance such as C_A' , Fig. 28.1, and R is determined chiefly by the grid-leak resistor R_A' , attached to the same grid. In fact, the simple expression

$$T = C_A'R_A' + C_B'R_B'$$

is a crude approximation for the duration of one complete cycle of the multivibrator (reciprocal of the natural frequency). It should

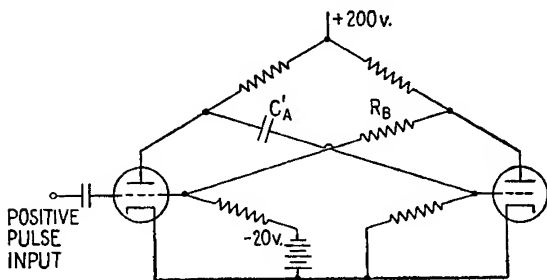


FIG. 28.4.—Hybrid of multivibrator and trigger circuit.

be emphasized, however, that this approximation gives the order of magnitude only and may be readily in error by as much as 50 per cent. Improved approximations are available though an exact solution is difficult because of the grossly nonlinear operation of the tube.

Multivibrators are relaxation oscillators, similar in general principle to the relaxation oscillator first discussed, Sec. 2, except that the critical potentials are determined by high-vacuum tubes rather than by gaseous devices. All previous remarks relative to synchronization and frequency reduction apply here. Though the multivibrator has a natural frequency of its own and will operate periodically without any assistance from outside sources, yet a fine adjustment of the operating frequency may be secured by synchronization. To this end a voltage having precisely the desired frequency or any low-order harmonic of the desired frequency is fed in to the oscillator circuit. The synchronizing signal then determines the exact *period*, while the multivibrator unit determines the *waveform* of the resultant cycle of operation. Many different

types of input coupling are in use. In fact, it would be difficult to inject a periodic voltage into a multivibrator circuit in any fashion without causing the device to "lock in" at some related frequency. For greatest sensitivity the synchronizing signal is usually injected in a grid circuit. A weak positive voltage can cause the grid to become unblocked slightly ahead of its natural uncontrolled schedule. If a weak negative voltage is present just before the crucial moment, the action is delayed slightly. Weak voltages received at other times in the cycle are disregarded, the grid being in an unresponsive condition.

If the circuit is to divide frequency by an even number, the synchronizing signal may be fed symmetrically to the two grids,

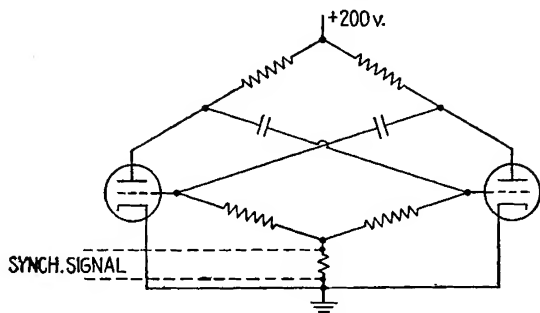


FIG. 28.5.—Injection of synchronizing signal for division by an even number.

Fig. 28.5. For example, in dividing by 6, the third cycle of the input wave may trip the left-hand grid, and then the sixth cycle of the input wave trips the right-hand grid in similar fashion. When dividing by an odd number, such as 5, the designer may first decide whether exact equality of the halves of the cycle is of any importance. If not, he may inject the synchronizing voltage on one grid only, using every fifth cycle to trip the left-hand tube. The length of the cycle as a whole is exactly correct; but the return stroke is unassisted, and the halves are not quite equal. Sometimes a deliberate inequality, such as 3 to 2, is planned for the two portions of the cycle. In this case the configuration of Fig. 28.5 may be retained, the signal is applied in similar fashion to each grid, but the time constants of the two grids differ in the correct proportion.

If equality of the halves is desired, in addition to correction of the cycle as a whole, then the odd-harmonic synchronizing signal should be fed to the two grids in opposite phase, Fig. 28.6. For example, input cycle $2\frac{1}{2}$ may assist the right-hand grid across its

cutoff threshold, cycle 5 renders similar service to the left-hand grid, cycle $7\frac{1}{2}$ again assists the right-hand grid, etc., for a net fivefold reduction of frequency.

Like the basic trigger circuits, multivibrators are often employed in cascade for the purpose of obtaining a large over-all reduction of frequency. Buffer amplifiers are often used as coupling devices between successive multivibrator pairs. Unlike the linear array of trigger circuits, the linear array of multivibrators is not limited to division by powers of 2 but may divide by any small whole number (say from 1 to 9) at each stage in the cascade circuit.

For example, assume that one has a sine wave or a series of pulses of frequency 13,230 cps and wishes to derive a frequency of

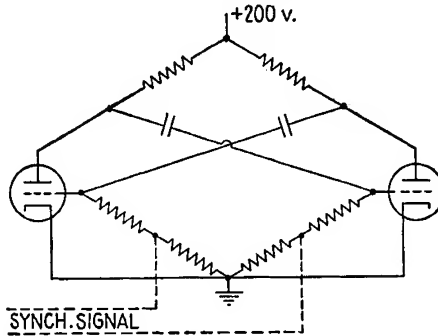


FIG. 28.6.—Injection of synchronizing signal for division by an odd number.

60 cycles, securely locked in a fixed phase relationship to the 13,230-cycle source. The desired reduction rate is $2/441$. The solution is obtained by first multiplying the source frequency by 2, using any conventional frequency-doubling device. The resulting 26,460-cycle voltage then is divided by 7. This is accomplished by injecting the 26,460-cycle synchronizing voltage on one grid of a multivibrator pair. This multivibrator is designed mathematically and if necessary adjusted experimentally so that its natural uncontrolled oscillation is in the neighborhood of 3,800 cps. Being considerably nearer to the seventh subharmonic of 26,460 than it is to the sixth or eighth subharmonic, application of moderate synchronizing voltage causes this multivibrator to snap into harmonic relationship at exactly 3,780 cps. (Excessive amplitude of synchronizing voltage would allow every sixth cycle to pull the trigger instead of waiting properly for the seventh. However, the available engineering margin is sufficient, and the excess amplitude can be adequately guarded against.)

The 3,780-cycle voltage just derived is passed through a buffer amplifier and then is used to synchronize the next multivibrator pair. This second pair resembles the first except that the coupling capacitances C_A' and C_B' , Fig. 28.1, are larger and the grid resistances R_A' and R_B' are also larger. The natural uncontrolled frequency as an independent relaxation oscillator is approximately 500 cps. Consequently, with proper amplitude of synchronizing voltage, the multivibrator snaps into synchronism at exactly one-seventh of 3,780 cycles, having shifted to precisely 540 cps. The further reduction to 60 cycles may be obtained in a single jump, reducing by a factor of 9, or it may be obtained in two stages, reducing by 3 and again by 3. Though more expensive, the latter choice offers a greater margin for security of the system as a whole against possible error due to voltage amplitude of drift. For this reason, it is adopted often in television control systems, where even a momentary error could cause considerable embarrassment and trouble. Reduction by larger jumps would be permissible in applications where less dependence need be placed upon security.

In addition to their very common use as elements in a frequency-reduction chain, multivibrators are also employed as frequency multipliers. For example a 50-kc multivibrator, accurately synchronized by a 50-kc crystal oscillator, may be used as the source of a series of harmonics, evenly spaced 50 kcps apart, and extending far into the high-frequency radio spectrum. This application results from the rich harmonic content that is indicated by the extremely angular waveform of the device. In ordinary multiplication the frequency ratios are integers but are not limited to small numbers. Ratios of 100 or more are obtainable rather easily. Such harmonic series are very useful in the calibration of receiving sets, wavemeters, signal generators, etc. Intermediate settings may be obtained by switching on a 10-kc multivibrator, synchronized by the same crystal oscillator. A further reduction by a factor of 10 (usually in two stages) permits generation of a precision 1,000-cycle wave. This may be used to drive an electric clock, verifying the average precision of the entire system. Additional verification is available by checking known high-frequency harmonics against the calibration signals radiated by the Bureau of Standards. Fractional frequency multiplication may be obtained by matching any known harmonic of a variable-frequency oscillator against a different known harmonic of the standard multivibrator.

Since the plate current is practically the same as the cathode

current, an approximately rectangular output voltage may be obtained from the multivibrator by using cathode-resistor output coupling. The missing upper right-hand corner of the rectangle, indicated by the dotted line at the top of Fig. 28.3, is sometimes negligible and frequently unobjectionable. Such voltage waves are useful in determining the response of amplifiers, in electronic switching of cathode-ray tube patterns, in driving differentiators as pulse makers, and in numerous other applications. Somewhat similar output voltage also may be derived by coupling to the plate of either tube, though here the capacitor-charging transients further modify the voltage pattern.

Unbalanced multivibrators with cathode-resistor output or with plate-circuit output can furnish a short, sharp, and approximately rectangular voltage pulse. Since this may be made to occupy any reasonable fraction of the cycle, it is particularly useful in the operation of keying tubes. In this application the unbalanced multivibrator is usually the final stage of a cascade frequency-reducing chain. The preceding stages lock the low-frequency keying pulse securely in any desired phase or time relationship with the higher frequency signal that is to be keyed in or keyed out.

Asymmetrical (unbalanced) multivibrators may do certain specialized jobs such as the production of a saw-tooth wave within the multivibrator circuit. Hybrid multivibrator triggers are useful in single-sweep and delayed-sweep applications such as recording the transients in a transmission line following a lightning stroke, observing double-refraction effects in a limited portion of an ionospheric sweep pattern, etc.

Increased precision of multivibrator frequency control may be secured by connecting one or both grid resistors to a moderately high positive-bias source instead of connecting the grid resistors to the cathode (as in the conventional multivibrator circuit). This causes the graph of the grid potential of a tube returning from its blocked condition to intersect the cutoff line at a steeper slope than would otherwise be obtained, thus improving the accuracy of timing. Minor frequency adjustments may be made by changing the positive bias.

APPENDIX A

REVIEW OF MATHEMATICS

I. EXPONENTIALS AND LOGARITHMS

1. Extension of the Notation of Exponents.—In the usual notation 2^3 means the product of three numbers each equal to 2, and in general a^m , where m is a positive integer, means the product of m numbers each equal to a . The fundamental rule holds:

$$a^m \cdot a^n = a^{m+n} \quad (1.1)$$

Successive extensions of the notation a^m now will be made such that rule (1.1) will always hold.

(1) a^{-m} means $\frac{1}{a^m}$; for example, $2^{-3} = \frac{1}{2^3} = \frac{1}{8}$

(2) a^0 means 1; for example, $2^0 = 1$, and $10^0 = 1$, . . .

(3) $a^{1/q}$ means $\sqrt[q]{a}$; for example, $3^{\frac{1}{3}}$ means

$$\sqrt[3]{3} = \sqrt{1.732} = 1.316$$

If q is an even number, a should be positive for $a^{1/q}$ to be real.

(4) $a^{p/q}$ means $(\sqrt[q]{a})^p$; for example, $2^{\frac{3}{2}}$ means

$$(\sqrt{2})^3 = (1.414)^3 = 2.828$$

Thus, if m is any rational number, positive, negative, zero, and a is a positive number, the meaning of a^m is clear.

2. Logarithms to the Base 2.—Consider the geometrical progression of ratio 2,

$$\cdots \frac{1}{8} \quad \frac{1}{4} \quad \frac{1}{2} \quad 1 \quad 2 \quad 4 \quad 8 \cdots \quad (2.1)$$

or

$$\cdots 2^{-3} \quad 2^{-2} \quad 2^{-1} \quad 2^0 \quad 2^1 \quad 2^2 \quad 2^3 \cdots \quad (2.2)$$

and the arithmetical progression of the exponents of 2,

$$\cdots -3 \quad -2 \quad -1 \quad 0 \quad 1 \quad 2 \quad 3 \cdots \quad (2.3)$$

3, say, is called the *logarithm of 8 to the base 2*. Any positive number, not only those of the progression (2.1), has a logarithm to the base 2. For example,

$$(2)^{\frac{3}{2}} = \sqrt[3]{32} = 3.174$$

hence,

$$\log_2 (3.174) = \frac{5}{3} = 1.667$$

In general,

$$\log_2 (x) = y \quad \text{means} \quad 2^y = x \quad (2.4)$$

3. Common or Decimal Logarithms.—Common or decimal logarithms are to the base 10; they are based on the geometrical progression

$$\cdots \frac{1}{1,000} \quad \frac{1}{100} \quad \frac{1}{10} \quad 1 \quad 10 \quad 100 \quad 1,000 \cdots$$

or

$$\cdots 10^{-3} \quad 10^{-2} \quad 10^{-1} \quad 10^0 \quad 10^1 \quad 10^2 \quad 10^3 \cdots$$

and the arithmetical progression

$$\cdots -3 \quad -2 \quad -1 \quad 0 \quad 1 \quad 2 \quad 3 \cdots$$

Thus -3 , say, is the common logarithm of 0.001 ; and, in general,

$$y = \log_{10} x \quad \text{means} \quad 10^y = x \quad (3.1)$$

4. Definition of the Number ϵ .—All curves $y = a^x$, where a is

a positive number, have the same general shape, Fig. 4.1. They all pass through the point $x = 0$, $y = 1$. For x positive they increase rapidly, the faster the larger a is; for x negative they approach zero rapidly, the faster the larger a is. Note that a^x is *always positive*. At point $(0,1)$ the slope of $y = 2^x$ (dashed line) is 0.693 , and the slope of $y = 10^x$ (solid line) is 2.30 . There is therefore a certain number ϵ , between 2 and 10 , such that the slope of $y = \epsilon^x$ at $(0,1)$ is unity. This important number is

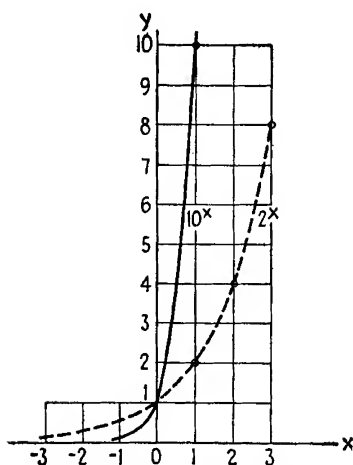


FIG. 4.1.—Graphs of the exponential functions 2^x and 10^x .

$$\epsilon = 2.71828 \cdots \quad (4.1)$$

5. Natural, or Napierian, Logarithms.—Logarithms to the base ϵ are called natural, or Napierian, logarithms.

$$\log_{\epsilon} x = y \quad \text{means} \quad \epsilon^y = x \quad (5.1)$$

The symbol $\ln x$ is sometimes used instead of $\log_{\epsilon} x$.

The graphs for $y = \log_{10} x$ and $y = \log_2 x$ are given in Fig. 5.1. Note that, since a^x is always positive, a negative number has no logarithm.

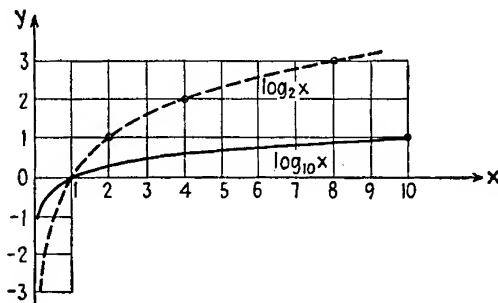


FIG. 5.1.—Graphs of the logarithmic functions $\log_2 x$ and $\log_{10} x$.

6. Use of Logarithms.—If $x_1 = \log_a y_1$, $x_2 = \log_a y_2$, then $y_1 = a^{x_1}$, $y_2 = a^{x_2}$, and

$$y_1 y_2 = a^{x_1} a^{x_2} = a^{(x_1 + x_2)} \quad (6.1)$$

also,

$$\frac{y_1}{y_2} = \frac{a^{x_1}}{a^{x_2}} = (a^{x_1})(a^{-x_2}) = a^{(x_1 - x_2)} \quad (6.2)$$

Rules: 1. Logarithm of product = sum of logarithms.

2. Logarithm of quotient = difference of logarithms.

This is true whatever the base a .

If a product consists of n factors each equal to x , then, from rule 1,

$$\log_a (x^n) = n \log_a x \quad (6.3)$$

Because of the generalizations in Sec. 1, this rule is equally true if n is any rational number, positive, negative, zero. (x must be positive in every case.)

7. To Find the Common Logarithm of a Number.—This given number must be positive; negative numbers have no logarithm.

1. If the given number N lies between 1 and 10: Say $N = 6.73$. Disregarding the decimal point, opposite the number 673 the logarithm table or slide rule gives 828. This will be the decimal part of the logarithm you seek. The logarithm must lie between 0 and 1; hence, $\log 6.73 = 0.828$.

2. If the given number N is greater than 10: Write it as the product of a power of 10 and a number larger than 1 and smaller than 10. Say $N = 673$, then $N = 10^2 \cdot 6.73$, and $\log 673 = 2.828$.

3. *If the given number N is smaller than unity:* Write it as the product of a negative power of 10 and a number larger than 1 and smaller than 10. Say $N = 0.0673$, then $N = 10^{-2} \cdot 6.73$, and $\log 0.0673 = -2 + 0.828$. This may also be written as $8.828 - 10$ or as -1.172 . Note that, in this last form, the decimal part of the logarithm does not consist of the figures given by the logarithm table or slide rule.

8. To Find the Natural Logarithm of a Number.—Here the decimal point must *not* be disregarded, because multiplying or dividing a given number by a power of 10 alters altogether its natural logarithm.

From a table of natural logarithms (or of the exponential function e^x , which amounts to the same thing), read the required number. For example,

$$\log_e 67.3 = 4.21 \quad \log_e 0.0673 = -2.70$$

If no table of natural logarithms is at hand, first find the common logarithm of the number, then multiply it by

$$\log_e 10 = 2.30258$$

For example,

$$\begin{array}{llll} \log_{10} 67.3 & = & 1.828, & \text{hence} \quad \log_e 67.3 & = & 4.21 \\ \log_{10} 0.0673 & = & -1.172, & \text{hence} \quad \log_e 0.0673 & = & -2.70 \end{array}$$

To prove the above rule, write

$$y = 10^{x_1} = e^{x_2}$$

Taking common logarithms,

$$x_1 = x_2 \log_{10} e = (x_2) 0.43429 \quad (8.1)$$

Taking natural logarithms,

$$x_2 = x_1 \log_e 10 = (x_1) 2.30258 \quad (8.2)$$

The following are convenient approximate values of powers of e (within 1 per cent):

$$\begin{array}{llllll} e^2 \doteq 7.4 & e^3 \doteq 20 & e^4 \doteq 55 & e^5 \doteq 150 & & \\ & & & e^6 \doteq 400 & e^7 \doteq 1,100 & \end{array}$$

9. Decreasing Exponential Function.—The function

$$y = 2^{-x} = \frac{1}{2^x} \quad (9.1)$$

can be represented by a graph, Fig. 9.1. This is the image of the graph of 2^x Fig. 4.1, reflected in the y axis.

An important example of such a function is the power along a nonresonant transmission line as a function of the distance from the generator. Because of losses along the line, the power at the end of a line 100 miles long may be only $\frac{1}{2}$, for example, of the input power. Then the power at the end of a line 200 miles long will be $\frac{1}{2}$ of this, or $\frac{1}{4}$ of the input power, etc. The plot of power against distance will be a decreasing exponential curve.

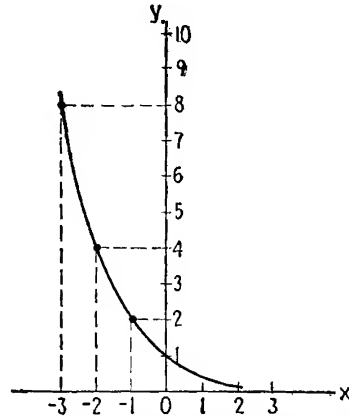


FIG. 9.1.—Decreasing exponential function $y = 2^{-x}$.

10. Power Ratios.—When the power along a telephone line is getting too low, it may be boosted by an amplifier. Consider a line OC , with amplifiers at A and B ; a diagram representing the variations in power along this line is given in Fig. 10.1. The input power P_i is 1 watt. The problem is to calculate the output power P_o at C , having the following data:

Ratio of power along the section of line OA , $P_i/P_1 = 12.6$.

Amplification of power by the amplifier at A , $P_1'/P_1 = 17.4$.

Ratio of power along the section of line AB , $P_1'/P_2 = 85.1$.

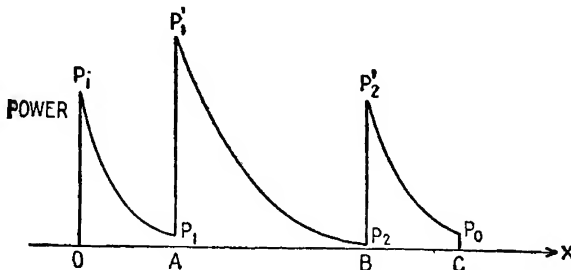


FIG. 10.1.—Power along a line with two amplifiers, one at A and the other at B .

Amplification of power by the amplifier at B , $P_2'/P_2 = 60.3$.

Ratio of power along the section of line BC , $P_2'/P_o = 9.77$.

The output power P_o is

$$P_o = P_i \frac{(17.4)(60.3)}{(12.6)(85.1)(9.77)}$$

This is a typical case for using logarithms (or a slide rule, on which the spacing of the numbers is determined by their logarithms). The logarithms to the base 10 are

$$\begin{array}{lll} \log 12.6 = 1.10 & \log 17.4 = 1.24 & \log 85.1 = 1.93 \\ \log 60.3 = 1.78 & \log 9.77 = 0.99 & \end{array}$$

Therefore,

$$\log \frac{P_3}{P_0} = -1.10 + 1.24 - 1.93 + 1.78 - 0.99 = -1.00$$

and

$$P_3 = 0.1P_0, \text{ or } 0.1 \text{ watt}$$

Thus to every section of line a *negative* number may be attached, and to every amplifier a *positive* number, which will be in every case the logarithm of the ratio of the output power to the input power. The magnitude of each negative number is called the *loss* of the specific section of line; each positive number is called the *gain* of

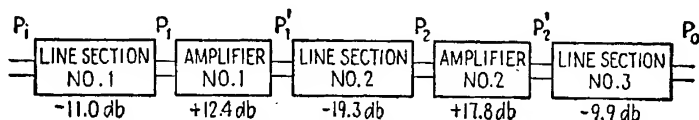


FIG. 10.2.—Block diagram of the line of Fig. 10.1.

the specific amplifier. To obtain the total gain or loss of the whole line *OC*, add all these numbers algebraically.

In this application, the common logarithms are said to be counted in *bels*, as, for example, -1.10 bel, 1.24 bel, etc. The bel being too large a unit, the practical unit is the *decibel*, which is 0.1 bel. In the above example the loss of the section of line *OA* is 11.0 db, the loss of section *AB* is 19.3 db, the loss of section *BC* is 9.9 db. The gain of the amplifier *A* is 12.4 db, and the gain of the amplifier *B* is 17.8 db. The block diagram of the whole line is given in Fig. 10.2. In general, the gain *G* in decibels is defined, for any network, by

$$\frac{P_o}{P_i} = 10^{0.1G} \quad \text{or} \quad G_{\text{db}} = 10 \log \frac{P_o}{P_i} \quad (10.1)$$

where P_i and P_o are the input and output power, respectively, of the network. In words, " P_o is G db higher than P_i " if G is positive or " G db lower than P_i " if G is negative. The diagram of Fig. 10.3, in which powers are plotted vertically on a logarithmic scale, illus-

trates this figurative way of speaking. It also shows that a terminal amplifier, with a gain of 10 db, would restore the original power level.

It is apparent that the decibel measures only power ratios and is therefore independent of the magnitude of the powers involved. It is possible to specify as zero level any convenient power level, in which case the magnitude of a power may be expressed in terms of decibels without ambiguity.

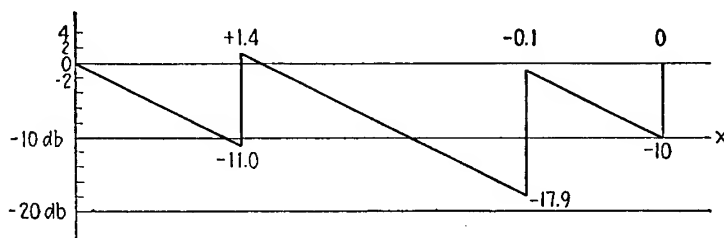


FIG. 10.3.—Power levels along a line on a logarithmic scale.

11. Voltage and Current Ratios.—Voltage ratios often have to be calculated. Let E_1 and E_2 be voltages, and assume that E_1 is applied to a resistor of R ohms and E_2 to another resistor also of R ohms. The currents in the resistors will be

$$I_1 = \frac{E_1}{R}, \quad I_2 = \frac{E_2}{R} \quad (11.1)$$

If P_1, P_2 are the powers dissipated,

$$\frac{P_2}{P_1} = \left(\frac{E_2}{E_1}\right)^2 = \left(\frac{I_2}{I_1}\right)^2 \quad (11.2)$$

According to rule (6.3),

$$\log \left(\frac{P_2}{P_1}\right) = 2 \log \left(\frac{E_2}{E_1}\right) = 2 \log \left(\frac{I_2}{I_1}\right)$$

therefore, from (10.1),

$$G_{ab} = 20 \log \left(\frac{E_2}{E_1}\right) = 20 \log \left(\frac{I_2}{I_1}\right) \quad (11.3)$$

Formulas (11.3) can be applied strictly *only* when the voltages E_1, E_2 are applied to resistances of the *same* value. However, they are also used, irrespective of the resistance values, when voltage ratios are the only quantities of importance, for example, in voltage amplifiers.

12. Nepers and Conversion Factors.—Natural, or napierian, logarithms are sometimes used for computing voltage, or current, ratios. Two voltages, or currents, are said to differ by one *neper* when one of them is ϵ times as large as the other ($\epsilon = 2.718$). If

$$\frac{E_2}{E_1} = \frac{I_2}{I_1} = \epsilon^N \quad (12.1)$$

then

$$N_{\text{nep}} = \ln \left(\frac{E_2}{E_1} \right) = \ln \left(\frac{I_2}{I_1} \right) \quad (12.2)$$

This neper is scarcely ever used except in calculating voltage, or current, ratios along transmission lines. If the voltage E_R at the receiving end of a nonresonant line is less than the voltage E_S at the sending end, and $E_R = E_S \epsilon^{-\alpha}$, the line is said to have an *attenuation* of α nepers.

Decibels and nepers are pure numerics based on power ratio and voltage ratio, respectively. When two voltages, or currents, are applied to the same resistance, there is a conversion rule relating decibels and nepers, as follows: Let D db correspond to N nepers. From (10.1) and (12.1),

$$\frac{P_2}{P_1} = (10)^{0.1D} = \left(\frac{E_2}{E_1} \right)^2 = \epsilon^{2N} \quad (12.3)$$

Taking common logarithms,

$$\begin{aligned} 0.1D &= 2N \log_{10} \epsilon = 0.8686N \\ D &= 8.686N \end{aligned} \quad (12.4)$$

The *neper* is therefore a larger unit than the *decibel*.

$$1 \text{ neper} = 8.686 \text{ db} \quad (12.5)$$

$$1 \text{ db} = 0.115 \text{ neper} \quad (12.6)$$

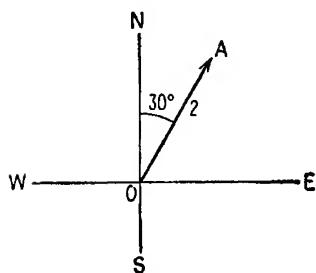


FIG. 13.1.—Physical quantities represented by a vector.

II. VECTORS AND COMPLEX ALGEBRA

13. Definition.—The vectors here considered serve as geometrical representation for any physical quantity that has both *magnitude* and *direction in a plane*. For example, displacement from a point O to a point A may be “2 miles” in the direction “North 30° East,” Fig. 13.1, or a force on a fixed point O may be “2 pounds” “upward, 30° to the right.” Lines such as OA , having length and

direction, can also be used to represent the magnitude and phase of alternating voltages and currents (Chap. I).

14. Addition of Vectors.—The vector sum of two vectors V_1 and V_2 is obtained by moving first along the vector V_1 , from O to A , Fig. 14.1, then moving from A to B , a distance equal to the magnitude of V_2 , in a direction parallel to V_2 . Calling V_{sum} the vector OB , $V_{\text{sum}} = V_1 + V_2$. The result would have been the same had the motion been first along V_2 , from O to C , then from C to B ,

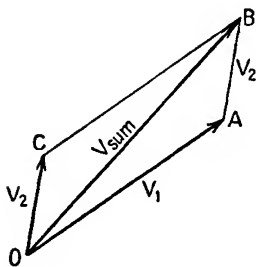


FIG. 14.1.—Addition of two vectors.

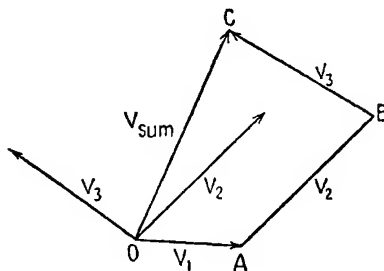


FIG. 14.2.—Addition of three vectors.

a “vector displacement” equal to V_1 , in magnitude and direction. Adding two vectors is the same as obtaining the resultant of two forces by the parallelogram rule.

In order to add three or more vectors V_1 , V_2 , V_3 , Fig. 14.2, move first from O to A along V_1 , then from A to B , a vector displacement equal to V_2 , then from B to C , a vector displacement equal to V_3 . The final result, the vector V_{sum} , or OC , is independent of the order in which the vectors V_1 , V_2 , V_3 are taken.

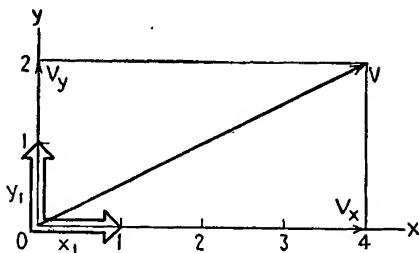


FIG. 15.1.—Components of a vector.

15. Components of a Vector.—A vector V is the sum of the two vectors V_x and V_y that are its projections on the usual rectangular axes Ox , Oy , Fig. 15.1.

The two vectors x_1 and y_1 of magnitude 1, directed in the positive

direction on Ox , Oy , are called *unit vectors*. For the example in Fig. 15.1,

$$V = 4x_1 + 2y_1 \quad (15.1)$$

which expresses that the displacement V is equivalent to moving four units to the right and two units upward. In general, if a and b are two numbers, positive, negative, zero, then, for any vector V in the plane,

$$V = ax_1 + by_1 \quad (15.2)$$

From the parallelogram rule, the sum of two vectors given by

$$V_1 = ax_1 + by_1 \quad V_2 = cx_1 + dy_1$$

is

$$V_3 = V_1 + V_2 = (a + c)x_1 + (b + d)y_1 \quad (15.3)$$

16. Complex Quantities.—The equation $x^2 + 1 = 0$ has no root, in the sense that no number x exists (positive, negative, zero) whose square is -1 . Adopting j as the symbol for $\sqrt{-1}$,

$$j^2 = -1 \quad (16.1)$$

the imaginary roots of $x^2 + 1 = 0$ are $\pm j$. Note that j is not a number.

An expression such as $a + jb$ is called a *complex quantity*. The *real part* is a ; the *imaginary part* is jb .

Complex numbers and in general an expression containing j can be treated by the ordinary rules of algebra; whenever j^2 occurs, replace it by -1 . For example,

$$2j(1 - j3)(2 + j) = 2j(2 - 5j - 3j^2) = 2j(5 - 5j) = 10 + j10$$

Complicated expressions containing j can be reduced, therefore, to the simple form $a + jb$.

17. Geometrical Representation of Complex Numbers.—Let z_1 and z_2 be two complex numbers.

$$z_1 = a + jb \quad z_2 = c + jd$$

Upon applying the ordinary rules of algebra, the sum z_3 of these two complex numbers is

$$z_3 = z_1 + z_2 = (a + c) + j(b + d) \quad (17.1)$$

The resemblance of (17.1) to (15.3) is striking; it has become common practice to represent geometrically the number 1 by the

horizontal unit vector x_1 and the symbol j by the vertical unit vector y_1 . A complex number $a + jb$ is represented geometrically by the vector $V = ax_1 + by_1$, Fig. 15.1. The sum z_3 of two complex numbers z_1 and z_2 is represented geometrically in Fig. 17.1.

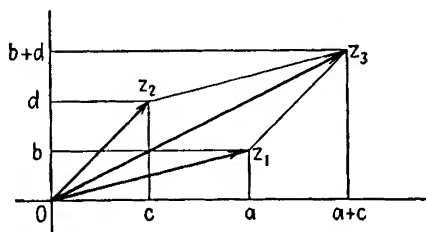


FIG. 17.1.—Addition of complex numbers.

18. Relations between the Two Definitions of a Complex Number.—A complex number and the vector that is its geometrical representation can be defined in two different ways,

1. By its rectangular components, a and jb .

2. By its magnitude M and the angle θ , measured positively (counterclockwise) from Ox .

In the first case,

$$z = a + jb \quad (18.1)$$

In the second case,

$$z = |z|/\theta \quad (18.2)$$

The complex number z is said to be represented in the *complex*, or *rectangular*, form (18.1) and in the *polar form* (18.2).

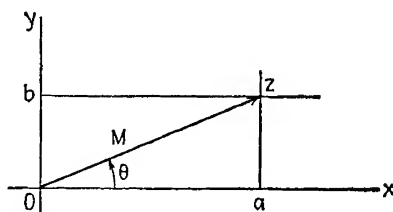
From Fig. 18.1, the relations

$$M = \sqrt{a^2 + b^2} \quad (18.3)$$

$$\cos \theta = \frac{a}{M} \quad (18.4)$$

$$\sin \theta = \frac{b}{M} \quad (18.5)$$

$$\tan \theta = \frac{b}{a} \quad (18.6)$$

FIG. 18.1.—Definition of a vector, or complex number, by (a, b) or by (M, θ) .

determine M and θ when a and b are given. If M and θ are given, a and b are determined by

$$a = M \cos \theta \quad (18.7)$$

$$b = M \sin \theta \quad (18.8)$$

19. Multiplication of Complex Numbers.—To obtain the product of two complex numbers, apply the ordinary rules of algebra, and

replace j^2 by -1 , (16.1). In the general case,

$$\begin{aligned} z_3 &= z_1 z_2 = (a + jb)(c + jd) \\ &= ac + jad + jbc + j^2 bd \\ &= (ac - bd) + j(ad + bc) \end{aligned} \quad (19.1)$$

The square of the magnitude of z_3 is

$$(ac - bd)^2 + (ad + bc)^2 = (a^2 + b^2)(c^2 + d^2) \quad (19.2)$$

and the angle of z_3 is given by

$$\begin{aligned} \tan \theta_3 &= \frac{ad + bc}{ac - bd} = \frac{(b/a) + (d/c)}{1 - (bd/ac)} \\ &= \frac{\tan \theta_1 + \tan \theta_2}{1 - \tan \theta_1 \tan \theta_2} = \tan (\theta_1 + \theta_2) \end{aligned} \quad (19.3)$$

In polar form (19.1) is equivalent to

$$(M_1/\theta_1)(M_2/\theta_2) = M_1 M_2 / \theta_1 + \theta_2 \quad (19.4)$$

or the magnitude of the product of two complex numbers is the *product of the magnitudes* of the components; the angle of the product is the *sum of the angles* of the components, Fig. 19.1.

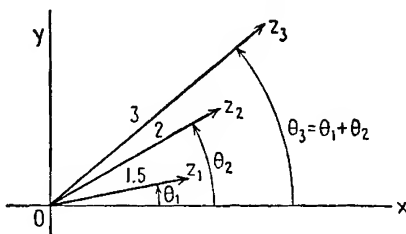


FIG. 19.1.—Product z_3 of two complex quantity z_1 and z_2 .

This algebraic result carries with it a rule for the multiplication of the two vectors V_1 and V_2 that represent geometrically z_1 and z_2 , viz., multiply the magnitudes of the two vectors, and add their angles, Fig. 19.1. Note that the “product of two vectors” thus obtained is neither the “scalar product ($A \cdot B$)” nor the “vector product ($A \times B$)” used in electromagnetic theory. The vectors that geometrically represent complex numbers and the vectors that represent the electromagnetic field are essentially different and obey different rules. The vectors that correspond to complex numbers are used in the vector diagrams of a-c theory. For example, in the generalized Ohm’s law $E = ZI$, if $Z = j10$, the

vector E will have ten times the magnitude of I and will lead I by 90° (counterclockwise).

If a complex quantity $z_1 = a + jb = 1/\theta$ has a magnitude 1 and an angle θ , multiplying any complex quantity z_2 by z_1 will have the effect geometrically of rotating the vector representing z_2 through the angle θ counterclockwise without changing its

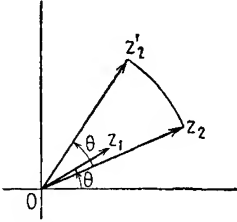


FIG. 19.2.—Product z_2' of a complex quantity z_2 and $z_1 = 1/\theta$.

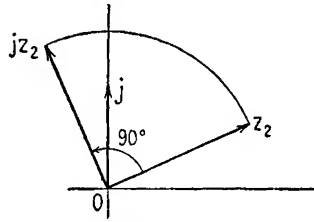


FIG. 19.3.—Product jz_2 of a complex quantity z_2 and j .

magnitude, Fig. 19.2. For this reason, complex quantities of magnitude 1 sometimes are called *rotators* or *phasors*.

In particular, multiplying a complex quantity by j has geometrically the effect of rotating it 90° counterclockwise, Fig. 19.3. Multiplying by $-j$ or $1/j$ has the effect of a clockwise rotation of 90° .

20. Conjugate Complex Numbers.—Consider the quadratic equation with real coefficients,

$$x^2 - 6x + 13 = 0 \quad (20.1)$$

The two roots of the quadratic

$$ax^2 + bx + c = 0 \quad (20.2)$$

are given by

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad (20.3)$$

Whenever the quantity $b^2 - 4ac$ is negative, the two roots will be complex numbers with the same real part and opposite imaginary parts; for example, in (20.1),

$$x = \frac{6 \pm \sqrt{-16}}{2} = 3 \pm j2 \quad (20.4)$$

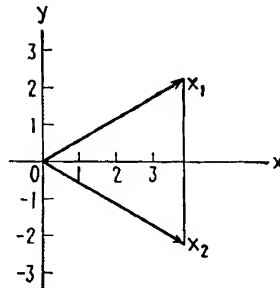


FIG. 20.1.—Example of conjugate complex numbers.

The representative vectors Ox_1 , Ox_2 in Fig. 20.1 are images of each other with respect to the Ox axis. Each of two such complex quantities is said to be the *conjugate* of the other.

The sum of two conjugate complex numbers $a \pm jb$ is always a real number $2a$; their difference is always a pure imaginary number $2jb$; their product is always a positive real number,

$$(a + jb)(a - jb) = a^2 - j^2b^2 = a^2 + b^2 \quad (20.5)$$

21. Ratio of Two Complex Numbers.—1. If the complex numbers z_1, z_2 are given in polar form,

$$\frac{M_1/\theta_1}{M_2/\theta_2} = \frac{M_1}{M_2} \frac{\theta_2}{\theta_1 - \theta_2} \quad (21.1)$$

i.e., magnitude of ratio = ratio of magnitudes; angle of ratio = algebraic difference of angles.

2. If z_1, z_2 are given in rectangular form, multiply numerator and denominator by the conjugate of the denominator. Thus,

$$\frac{a + jb}{c + jd} = \frac{(a + jb)(c - jd)}{(c + jd)(c - jd)} = \frac{ac + bd}{c^2 + d^2} + j \frac{bc - ad}{c^2 + d^2} \quad (21.2)$$

22. Square Root of Complex Number.—If the complex number z is given in rectangular form, first express z in polar form by the method of Sec. 18. Let $z = |z|/\theta$; then,

$$\sqrt{z} = \sqrt{|z|/\theta} = \sqrt{|z|} \left/ \frac{\theta}{2} \right. \text{ and also } = \sqrt{|z|} \left/ \frac{\theta}{2} + 180^\circ \right. \quad (22.1)$$

Magnitude of square root = square root of magnitude of original complex number; angle of square root = one-half the angle of original complex number and also = one-half this angle plus 180° , Fig. 22.1. This gives two opposite values for the square root, just as $\sqrt{4} = \pm 2$.

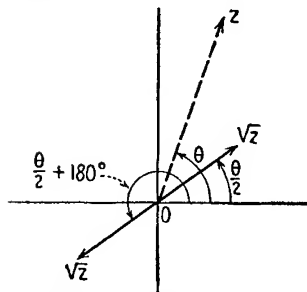


FIG. 22.1.—Square root of a complex number.

III. DIFFERENTIATION

23. Derivative.—If a quantity x varies with time, the rate of change of this quantity with respect to time is measured by $\tan \alpha$, the slope of the tangent to the graph of the function $x(t)$, Fig. 23.1. It is practically equal to the ratio $\Delta x/\Delta t$, where Δt is a small interval of time, and rigorously is the limit of $\Delta x/\Delta t$, when Δt diminishes to zero, provided that there is such a limit. It is denoted by

$$\tan \alpha \quad \text{or} \quad \frac{dx}{dt} \quad \text{or} \quad x'(t) \quad \text{or} \quad D_t x \quad (23.1)$$

As a function of t , dx/dt is called the *derivative* of the function $x(t)$. dx and dt (variable quantities approaching zero) are called the differentials of x and of t .

Example.—Let y be a function of x , say $y = x^2$ (parabolic). To find the derivative of $y(x)$ at point $x = x_0$, $y = y_0 = x_0^2$,

$$\begin{aligned} y_0 + \Delta y &= (x_0 + \Delta x)^2 \\ &= x_0^2 + 2x_0 \Delta x + (\Delta x)^2 \\ \Delta y &= (2x_0 + \Delta x) \Delta x \\ \frac{\Delta y}{\Delta x} &= 2x_0 + \Delta x \end{aligned}$$

The slope of a straight line MM' , Fig. 23.2, passing through the point M and cutting the parabola at another point M' to the right of M , is $2x_0$ plus a

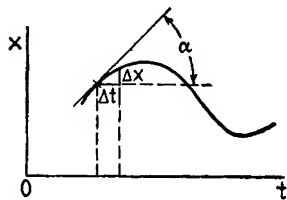


FIG. 23.1.—Derivative of a function.

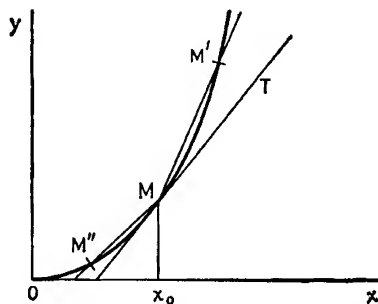


FIG. 23.2.—Illustrating the derivative of $y = x^2$.

positive quantity Δx . The slope of a straight line cutting the parabola at M and at another point M'' to the left of M is $2x_0$ plus a negative quantity $(-\Delta x)$. The slope of the tangent at M , that is, of a straight line passing through M and cutting the parabola neither to the right nor to the left, is exactly $2x_0$. Thus, the derivative of x^2 is $2x$.

24. Derivatives of Important Functions.—By this process the derivatives of the following important functions are found (in the following, a is a constant, x is a variable):

Function	Derivative	
a	0	(24.1)
x	1	(24.2)
x^2	$2x$	(24.3)
x^n	nx^{n-1}	(24.4)

$$\sqrt{x} \qquad \frac{1}{2\sqrt{x}} \qquad (24.5)$$

$$\epsilon^x \qquad \epsilon^x \qquad (24.6)$$

$$a^x \qquad (\log_e a)(a^x) \qquad (24.7)^1$$

$$\log_e x \qquad \frac{1}{x} \qquad (24.8)$$

$$\sin x \qquad \cos x \qquad (24.9)$$

$$\cos x \qquad -\sin x \qquad (24.10)$$

$$\tan x \qquad \frac{1}{\cos^2 x} = 1 + \tan^2 x \qquad (24.11)$$

$$\sin^{-1} x \qquad \frac{1}{\sqrt{1-x^2}} \qquad (24.12)$$

$$\cos^{-1} x \qquad \frac{-1}{\sqrt{1-x^2}} \qquad (24.13)$$

$$\tan^{-1} x \qquad \frac{1}{1+x^2} \qquad (24.14)$$

25. Useful Rules for Differentiation.—The process of obtaining the derivative is called differentiation. The derivatives of many other functions may be obtained from derivatives of the functions of Sec. 24 by applying the following rules in which u, v denote functions of x , and a is a constant:

$$\frac{d(u+v)}{dx} = \frac{du}{dx} + \frac{dv}{dx} \qquad (25.1)$$

$$\frac{d(au)}{dx} = a \frac{du}{dx} \qquad (25.2)$$

$$\frac{d(uv)}{dx} = u \frac{dv}{dx} + v \frac{du}{dx} \qquad (25.3)$$

$$\frac{d\left(\frac{u}{v}\right)}{dx} = \frac{v \frac{du}{dx} - u \frac{dv}{dx}}{v^2} \qquad (25.4)$$

$$\frac{d(\epsilon^u)}{dx} = \epsilon^u \frac{du}{dx} \qquad (25.5)$$

$$\frac{d(\log_e u)}{dx} = \frac{1}{u} \frac{du}{dx} \qquad (25.6)$$

$$\frac{d[f(u)]}{dx} = \frac{df(u)}{du} \frac{du}{dx} \qquad (25.7)$$

¹ Note from (24.6) and (24.7) that the number ϵ is such that the derivative of ϵ^x is the function itself, which is not true for any other positive number (such as 2 or 10) raised to the power x . This justifies the way in which ϵ was introduced in Sec. 4.

26. Examples.—The useful rule (25.7) of which (25.5) and (25.6) are particular cases, gives the derivative of a *function of a function of x* ; that is, of $f(u)$, where $f(u)$ is a function of u and u is a function of x . For example, to obtain the derivative of $\cos \omega x$, put $u = \omega x$; then

$$\frac{d(\cos \omega x)}{dx} = \frac{d(\cos u)}{du} \frac{du}{dx} = -\sin u \omega \quad (26.1)$$

Example of application of rule (25.3): To find the derivative of $y = \epsilon^{-ax} \cos \omega x$, put $\epsilon^{-ax} = u$, $\cos \omega x = v$,

$$\begin{aligned} \frac{d(\epsilon^{-ax} \cos \omega x)}{dx} &= \frac{d(uv)}{dx} = u \frac{dv}{dx} + v \frac{du}{dx} \\ &= \epsilon^{-ax}(-\omega \sin \omega x) + \cos \omega x(-a\epsilon^{-ax}) \\ &= -\epsilon^{-ax}(\omega \sin \omega x + a \cos \omega x) \end{aligned} \quad (26.2)$$

27. Maxima and Minima.—Since from (23.1) the derivative of a function is equal to the slope, $\tan \alpha$, of its graph, it is plain that wherever a function has a flat maximum or minimum the derivative of this function must be zero. For example, $\sin x$ is maximum for $x = 90^\circ, 450^\circ, \dots$, minimum for $x = -90^\circ, 270^\circ, \dots$. For these particular values of x , $\cos x$ is zero.

However, the derivative may be zero and the function have neither a maximum nor a minimum. For example, $y = x^3$ for $x = 0$. A general method for finding where a function is maximum or minimum is therefore to find the values of the variable for which the derivative is zero and to check what the function does for these values. For example, a quadratic function $y = ax^2 + bx + c$ has for its derivative $y' = 2ax + b$. For $x = -b/2a$, $y' = 0$, and y is maximum if $a < 0$, minimum if $a > 0$.

28. Derivatives of Higher Order.—The derivative of a function $y = f(x)$ is again a function of x , which in turn can be differentiated with respect to x . The *second derivative* of $f(x)$ is

$$\frac{d}{dx} \frac{dy}{dx} \quad \text{or} \quad \frac{d^2y}{dx^2} \quad \text{or} \quad f''(x) \quad \text{or} \quad D_{xx}x \quad (28.1)$$

The derivative of (28.1) is the *third derivative* of $f(x)$.

$$\frac{d}{dx} \frac{d^2y}{dx^2} \quad \text{or} \quad \frac{d^3y}{dx^3} \quad \text{or} \quad f'''(x) \quad \text{or} \quad D_{xxx}x \quad (28.2)$$

etc.

For example, if

$$\left. \begin{array}{l} \text{then} \\ y = \sin x \\ f'(y) = \cos x \\ f''(y) = -\sin x \\ f'''(y) = -\cos x \\ f''''(y) = \sin x \end{array} \right\} \quad (28.3)$$

Also, if

$$\left. \begin{array}{l} \text{then} \\ z = \cos x \\ f'(z) = -\sin x \\ f''(z) = -\cos x \\ f'''(z) = \sin x \\ f''''(z) = \cos x \end{array} \right\} \quad (28.4)$$

After four differentiations the original function returns. This is true only of the two functions $\sin x$ and $\cos x$ and linear combinations thereof, as $A \sin x + B \cos x$.

IV. INTEGRATION

29. Definition.—Integration is the reverse process of differentiation. For example, $2x$ is the derivative of x^2 ; hence x^2 is the *indefinite integral*, or *primitive*, of $2x$. Since the derivative of a constant is zero, then $x^2 + C$, where C is a constant, is also the indefinite integral of $2x$. The *integration constant* C is from this point of view undetermined, hence the name “indefinite integral.”

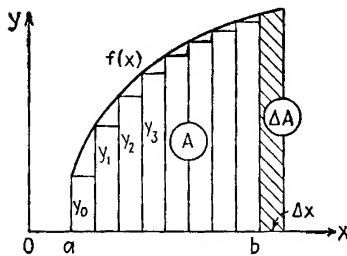


FIG. 30.1.—Definite integral as an area.

30. Definite Integral as Area. Consider the area A bounded by the graph of $y = f(x)$, the axis Ox , the ordinates $x = a$ and $x = b$, Fig. 30.1. Dividing the interval $a-b$ in a great number of small intervals of length Δx , and calling

y_0, y_1, y_2, \dots the successive ordinates of the curve, an approximate value for A is the sum of all the rectangles,

$$A \doteq y_0 \Delta x + y_1 \Delta x + y_2 \Delta x + y_3 \Delta x + \dots$$

The limit of this sum when $\Delta x \rightarrow 0$ is the area A .

$$A = \lim_{\Delta x \rightarrow 0} (y_0 \Delta x + y_1 \Delta x + \dots), \text{ } x \text{ varying from } a \text{ to } b. \text{ This}$$

is written as

$$A = \int_a^b y \, dx = \int_a^b f(x) \, dx \quad (30.1)$$

and is called a *definite integral*. The integral sign \int was originally S and stands for "sum" or, rather, "limit of sum."

31. Definite Integral Related to Primitive Function.—If x is increased by Δx and A by ΔA , Fig. 30.1, an approximate value of ΔA is $y_b \Delta x$, where y_b is the ordinate $f(b)$ for $x = b$. Then,

$$\frac{\Delta A}{\Delta x} \doteq y_b$$

and

$$\lim_{\Delta x \rightarrow 0} \left(\frac{\Delta A}{\Delta x} \right) = \left(\frac{dA}{dx} \right)_{x=b} = y_b = f(b) \quad (31.1)$$

Since the derivative of $A(x)$ is y or $f(x)$, $A(x)$ is the indefinite integral of $f(x)$; $A(x) = F(x) + C$, where $F(x)$ is a primitive function of $f(x)$. Now

$$A = \int_a^b y \, dx = \int_a^b \frac{dA}{dx} \, dx = \int_a^b dA = A_b - A_a \quad (31.2)$$

or the area A between the ordinates $x = a$ and $x = b$ is the difference of the values of $A(x)$ for $x = b$ and $x = a$,

$$A = [F(b) + C] - [F(a) + C] = F(b) - F(a) \quad (31.3)$$

In this expression for the area A by the integral $\int_a^b f(x) \, dx$, the integration constant C has disappeared and hence the name "definite integral."

32. Example.—The straight line OP , Fig. 32.1, has for its equation $y = mx$. To find the area of the triangle OQP .

$$\text{Primitive of } mx = \frac{mx^2}{2} + C$$

$$\begin{aligned} \text{Area } OPQ &= \left(\frac{mx^2}{2} + C \right)_{x=a} \\ &\quad - \left(\frac{mx^2}{2} + C \right)_{x=0} = \frac{ma^2}{2} \end{aligned} \quad (32.1)$$

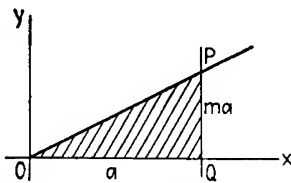


FIG. 32.1.—Integral of the function $y = mx$.

which agrees with the usual rule $\frac{1}{2}bh$ ($b = a$, $h = ma$).

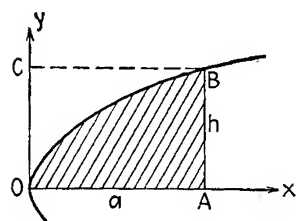
If y is the voltage E across a capacitor of capacitance C , $E = Q/C$. The energy stored in the capacitor equals

$$\int_0^Q E \, dQ = \frac{1}{2} \frac{Q^2}{C} \quad (32.2)$$

an important result.

33. Second Example.—Given the parabola $y = \sqrt{2px}$, to find the area limited by the arc OB of the parabola, the axis $y = 0$ and the ordinate AB ($OA = a$), Fig. 33.1.

The derivative of x^n is nx^{n-1} ; therefore the derivative of $x\sqrt{x}$ or $x^{\frac{3}{2}}$ is $\frac{3}{2}x^{\frac{1}{2}}$ or $\frac{3}{2}\sqrt{x}$; hence the indefinite integral of $\sqrt{2px}$ is $\frac{2}{3}\sqrt{2p}x^{\frac{3}{2}} + C$. Therefore,



$$\begin{aligned} \text{Area } OAB &= \left[\frac{2}{3} \sqrt{2p} x^{\frac{3}{2}} + C \right]_{x=0}^{x=a} \\ &= \frac{2}{3} \sqrt{2p} a \sqrt{a} \\ &= \frac{2}{3} \sqrt{2pa} a = \frac{2}{3} ha \quad (33.1) \end{aligned}$$

FIG. 33.1.—Area of sector of a parabola.

In words, “area OAB is two-thirds of the rectangle $OABC$.”

34. Indefinite Integrals of Important Functions.—These are obtained very simply from the list of derivatives, Sec. 24.

Function	Indefinite Integral	
1	$x + C$	(34.1)
x	$\frac{x^2}{2} + C$	(34.2)
x^n	$\frac{x^{n+1}}{n+1} + C$	(34.3)
$\frac{1}{\sqrt{x}}$	$2\sqrt{x} + C$	(34.4)
\sqrt{x}	$\frac{2}{3}x\sqrt{x} + C$	(34.5)
$\frac{1}{x}$	$\log_e x + C$	(34.6)
e^x	$e^x + C$	(34.7)
e^{ax}	$\frac{1}{a}e^{ax} + C$	(34.8)
$\cos x$	$\sin x + C$	(34.9)
$\sin x$	$-\cos x + C$	(34.10)
$\cos \omega x$	$\frac{1}{\omega} \sin \omega x + C$	(34.11)
$\sin \omega x$	$-\frac{1}{\omega} \cos \omega x + C$	(34.12)

V. SERIES DEVELOPMENTS AND APPROXIMATION FORMULAS

35. Polynomial in x ; x Large.—Consider a polynomial in x ,

$$y = 5 + 3x - 2x^2 + x^3 \quad (35.1)$$

The values of y for $x = 1, 10, 100, 1,000$ are

$$7 \quad 835 \quad 980,305 \quad 998,003,005 \quad (35.2)$$

For $x = 10,000$,

$$y = 999,800,030,005$$

If y is a physical quantity obtained from laboratory measurements and if the precision of the measurements is of the order of 0.1 per cent, only three significant figures should be retained in the numbers and the following figures should be discarded as meaningless. The values of y , (35.2), should be written as

$$7.00 \quad 835 \quad 980 \cdot 10^3 \quad 998 \cdot 10^6 \quad 1,000 \cdot 10^9$$

This last value is that of x^3 , and it is seen that in this case terms of lower degree than x^3 contribute nothing to the result.

When x is sufficiently large, a polynomial in x is *approximately equal* to its term of highest degree, or *equivalent* to it; in symbols,

$$y \approx x^3 \quad \text{or} \quad y \doteq x^3 \quad (35.3)$$

when x is very large with respect to 1 (indicated by $x \gg 1$).

36. Polynomial in x ; x Small.—Again consider the polynomial of Sec. 35,

$$y = 5 + 3x - 2x^2 + x^3 \quad (36.1)$$

The values of y for $x = 1, 0.1, 0.01, 0.001$ are

$$7 \quad 5.281 \quad 5.029801 \quad 5.002998001$$

and, for $x = 0$,

$$y = 5.000 \dots$$

If the precision of the measurements is of the order of 0.1 per cent, these numbers should be written

$$7.00 \quad 5.28 \quad 5.03 \quad 5.00 \quad 5.00$$

All terms of the polynomial other than the first contribute nothing to the last two results.

When x is sufficiently small, a polynomial in x is approximately equal to its term of lowest degree; in symbols,

$$y \approx 5 \quad \text{or} \quad y \doteq 5 \quad (36.2)$$

when x is very small with respect to 1 (indicated by $x \ll 1$).

37. Series Developments.—Consider the expression $1/(1 - x)$; since

$$1 = (1 - x)(1 + x + x^2 + x^3 + \dots) \quad (37.1)$$

then

$$y = \frac{1}{1 - x} = 1 + x + x^2 + x^3 + \dots \quad (37.2)$$

The expression at the right of (37.2), which contains an infinite number of terms, is the *series development* of $1/(1 - x)$. It is valid whenever the absolute value of x is less than unity. If, for example, $x = 0.1$,

$$\frac{1}{1 - 0.1} = 1 + 0.1 + 0.01 + 0.001 + 0.0001 + \dots \quad (37.3)$$

which is merely the usual decimal expansion of $\frac{10}{9} = 1.1111 \dots$

Suppose x and y are physical quantities, measured with a precision of 0.1 per cent. For $x = 0.1, 0.01, 0.001$, the values of y are

$$1.1111 \dots \quad 1.010101 \dots \quad 1.001001001 \dots$$

or, retaining only three significant figures,

$$1.11 \quad 1.01 \quad 1.00$$

If x is small enough, the infinite series development (37.2) is equivalent to its first term (the same rule as for a polynomial). However, if x , while small, is large enough to be retained in the expression $1 - x$, as for example $x = 0.01$, it should also be retained in the series development, but x^2 and higher powers may be discarded. Then

$$\frac{1}{1 - x} \doteq 1 + x \quad \text{when } x \ll 1 \quad (37.4)$$

This is an *approximation formula* of frequent use.

Also,

$$\frac{1}{1 + x} \doteq 1 - x \quad \text{when } x \ll 1 \quad (37.5)$$

If x were larger, say $x = 0.1$, then x and x^2 should be retained, x^3 and higher powers discarded.

38. Binomial Theorem.—The following development is called the binomial theorem:

$$\begin{aligned}
 (a + b)^n &= a^n + na^{n-1}b + \frac{n(n-1)}{1 \cdot 2} a^{n-2}b^2 \\
 &\quad + \frac{n(n-1)(n-2)}{1 \cdot 2 \cdot 3} a^{n-3}b^3 + \dots \\
 &\quad + \frac{n(n-1) \dots (n-p+1)}{p!} a^{n-p}b^p + \dots \quad (38.1)
 \end{aligned}$$

where $p!$ or $p|$ (factorial p) stands for the product $(1 \cdot 2 \cdot 3 \dots p)$.

If n is a positive integer, the formula contains only $n + 1$ terms. Thus,

$$(a + b)^2 = a^2 + 2ab + b^2 \quad (38.2)$$

$$(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3 \quad (38.3)$$

etc.

If n is a positive or negative real number but not a positive integer, the right-hand member contains an infinite number of terms and is the *series development* of $(a + b)^n$. Examples:

For $n = -1$,

$$\begin{aligned}
 (a + b)^{-1} &= a^{-1} + (-1)a^{-2}b + \frac{(-1)(-2)}{1 \cdot 2} a^{-3}b^2 \\
 &\quad + \frac{(-1)(-2)(-3)}{1 \cdot 2 \cdot 3} a^{-4}b^3 + \dots \\
 &= a^{-1} - a^{-2}b + a^{-3}b^2 - a^{-4}b^3 + \dots \quad (38.4)
 \end{aligned}$$

For $n = \frac{1}{2}$,

$$\begin{aligned}
 (a + b)^{\frac{1}{2}} &= a^{\frac{1}{2}} + \frac{1}{2}a^{-\frac{1}{2}}b + \frac{\frac{1}{2}(-\frac{1}{2})}{1 \cdot 2} a^{-\frac{3}{2}}b^2 \\
 &\quad + \frac{\frac{1}{2}(-\frac{1}{2})(-\frac{3}{2})}{1 \cdot 2 \cdot 3} a^{-\frac{5}{2}}b^3 + \dots \\
 &= a^{\frac{1}{2}} + \frac{1}{2}a^{-\frac{1}{2}}b - \frac{1}{8}a^{-\frac{3}{2}}b^2 + \frac{3}{128}a^{-\frac{5}{2}}b^3 - \dots \quad (38.5)
 \end{aligned}$$

This development is valid whenever the absolute value of (b/a) is less than unity.

39. Square Root of Numbers Near Unity.—Replace a by 1 and b by x in (38.5). Then,

$$\sqrt{1 + x} = 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \frac{3}{128}x^3 - \dots \quad (39.1)$$

which is the series development of $\sqrt{1 + x}$. If x is very small with regard to 1, the very useful approximation formula (39.2) results,

$$\sqrt{1 + x} \doteq 1 + \frac{x}{2} \quad \text{when} \quad x \ll 1 \quad (39.2)$$

Examples:

$$\sqrt{1.004} \doteq 1.002 \quad \sqrt{3,606} \doteq 60 \left(1 + \frac{3}{3,600} \right) \doteq 60.05$$

40. Taylor's Series.—The above series are particular cases of the very general *Taylor's series development*. Let curve C , Fig.

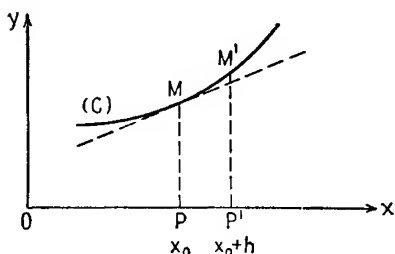


FIG. 40.1.—Illustrating Taylor's series.

40.1, be the graph of $y = f(x)$. Taylor's formula states that, starting at point M on the curve, the value of $f(x)$ can be found at a neighboring value of x , say $x_0 + h$, provided that the values of the function $f(x)$ and of its first and subsequent derivatives are known for

$x = x_0$. The more of these derivatives which are known, the more precise will be the result for a given value of h .

Taylor's formula is

$$f(x_0 + h) = f(x_0) + f'(x_0)h + f''(x_0) \frac{h^2}{2!} + f'''(x_0) \frac{h^3}{3!} + \cdots + R \quad (40.1)$$

$f'(x_0)$ is the slope of the tangent to C at point M . Ending Taylor's formula with its second term is equivalent to simulating curve C by its tangent at M . In so doing an error is committed, and in rigorous work an upper limit of this error should be known. The following formula is *exact*:

$$f(x_0 + h) = f(x_0) + f'(x_0)h + \underbrace{f''(x_0 + \theta h) \frac{h^2}{2}}_R \quad (40.2)$$

The error, or "remainder" R , is equal to $f''(x_0 + \theta h) \frac{h^2}{2}$, where θ lies between zero and unity, but is otherwise unknown. When θ varies from 0 to 1, the point of abscissa $(x_0 + \theta h)$ travels from M to M' on curve C ; therefore, if a positive number K can be found such that $|f''(x)|$ is always smaller than K on arc MM' , the error in using the approximation formula

$$f(x_0 + h_2) \doteq f(x_0) + f'(x_0)h \quad (40.3)$$

will be smaller than $Kh^2/2$.

$f''(x_0)$ is the rate of change of the slope of curve C , taken at point M . If $f''(x_0)$ is positive, curve C gets steeper when M moves to the right, or C is concave upward (as in Fig. 40.1). If $f''(x_0)$ is negative, C is concave downward. Taylor's formula ended with its third term is

$$f(x_0 + h) = f(x_0) + f'(x_0)h + f''(x_0)\frac{h^2}{2} + \underbrace{f'''(x_0 + \theta h)\frac{h^3}{6}}_R \quad (40.4)$$

If curve C represented the position of a car on a road as a function of time $x = f(t)$ and the axes were marked (Ot, Ox) instead of (Ox, Oy), $f'(t_0)$ would be the speed of the car at time t_0 , and $f''(t_0)$ its acceleration at the same instant.

$f'''(x_0)$ is the rate of change of $f''(x)$ at point $x = x_0$; if it is positive, curve C is flatter at the left of M and is more curved at the right of M . The higher derivatives have no obvious geometrical meaning.

The detector action of a diode depends essentially on the $f''(x)$ of its characteristic. The determination of the current amplitude in an oscillator depends upon $f'''(x)$. Higher derivatives are seldom needed in radio.¹

41. Maclaurin's Series.—When $x_0 = 0$, Taylor's series has a slightly simplified form. Writing x for h ,

$$f(x) = f(0) + f'(0)x + f''(0)\frac{x^2}{2} + f'''(0)\frac{x^3}{6} + \cdots + R \quad (41.1)$$

which is called *Maclaurin's series* and is frequently used.

42. Useful Series Developments.—The derivative of e^x is e^x , (24.6), and all the successive derivatives of e^x are e^x . For $x = 0$ all are 1, and Maclaurin's formula for e^x is

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \quad (42.1)$$

which is the series development of e^x .

For $x = 0$ the successive derivatives of $\sin x$ take the values 1, 0, -1, 0, 1, 0, \cdots . Maclaurin's formula for $\sin x$ is therefore

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots \quad (42.2)$$

¹ See, however, Chaffee, "Theory of Thermionic Vacuum Tubes," p. 483, McGraw-Hill Book Company, Inc., 1933.

(every other term is missing in this development). Similarly,

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots \quad (42.3)$$

These three developments are valid for all values of x , large or small.

These developments give the following useful *approximation formulas* when $x < 1$:

$$e^x \doteq 1 + x \quad (42.4)$$

$$\sin x \doteq x \quad (42.5)$$

$$\cos x \doteq 1 - \frac{x^2}{2} \quad (42.6)$$

$$\tan x \doteq x \quad (42.7)$$

43. Approximation Rules for Impedance.—An impedance Z is the complex sum of a resistance and a reactance; $Z = R + jX$ (Chap. I). Its magnitude is $|Z| = \sqrt{R^2 + X^2}$. Hence,

If

$$R > 10X, \quad |Z| \doteq R \quad (43.1)$$

If

$$|X| > 10R, \quad |Z| \doteq |X| \quad (43.2)$$

to an approximation better than 0.5 per cent, (39.2).

The angle of Z is given by $\tan \theta = X/R$; hence, from (42.7),

If

$$R > 10X, \quad \theta \doteq \frac{X}{R} \text{ radians} \quad (43.3)$$

If

$$|X| > 10R, \quad |\theta| = \frac{\pi}{2} - \phi \quad (43.4)$$

where $\phi \doteq R/|X|$ radians. The angle ϕ is the complement of the angle θ .

If the small reactance of Z , neglected in (43.1), is desired, it may be found from

$$X = R \tan \theta \doteq R\theta \quad (43.5)$$

with θ in radians. Similarly, the small resistance of Z , neglected in (43.2), is found from

$$R = \frac{X}{\tan \theta} = |X| \tan \phi \doteq |X|\phi \quad (43.6)$$

where ϕ is the complement of $|\theta|$.

These approximations are very important in practice. For example, if a capacitor C is shunted by a resistor R , the combination will behave practically as a resistor for frequencies so low that

$$\frac{1}{\omega C} > 10R \quad \text{or} \quad \omega < \frac{1}{10CR}$$

and practically as a capacitor for frequencies so high that

$$R > \frac{10}{\omega C} \quad \text{or} \quad \omega > \frac{10}{CR}$$

44. Euler's Identity.—It can be shown that the developments of Sec. 42 are valid when x takes any constant value, real or complex. This being assumed, replace x by $j\theta$ in (42.1), and

$$\begin{aligned} e^{j\theta} &= 1 + j\theta + j^2 \frac{\theta^2}{2!} + j^3 \frac{\theta^3}{3!} + \cdots \\ &= \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots\right) + j\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots\right) \end{aligned} \quad (44.1)$$

The real part of $e^{j\theta}$ is the series development (42.3) of $\cos \theta$, and the imaginary part is the development (42.2) of $\sin \theta$. Equation (44.1) takes the remarkable form

$$e^{j\theta} = \cos \theta + j \sin \theta \quad (44.2)$$

sometimes called *Euler's identity*, Fig. 44.1.

Since $\cos^2 \theta + \sin^2 \theta = 1$, the complex number $e^{j\theta}$ has unity for modulus and θ for argument. For this reason, $e^{j\theta}$ is sometimes called a *rotator* or *phasor*, Sec. 19.

Any complex number, therefore, may be expressed by the equivalent forms

$$z = a + jb = \rho \cos \theta + j\rho \sin \theta = \rho e^{j\theta} \quad (44.3)$$

In the last expression the angle θ of the complex number appears in the exponent of the napierian base e . The magnitude ρ of z may also be expressed as a power of e , as $\rho = e^m$. Then

$$z = a + jb = \rho e^{j\theta} = e^m e^{j\theta} = e^{m+j\theta} \quad (44.4)$$

In communication theory a complex quantity similar to that in

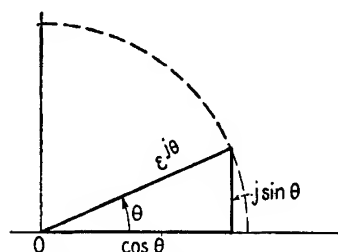


FIG. 44.1.—Illustration of Euler's identity.

(44.4) is often encountered. This quantity is

$$\epsilon^{-\alpha-j\beta} = \epsilon^{-\alpha}\epsilon^{-j\beta} \quad (44.5)$$

where α is usually positive, while β may be either positive or negative. The magnitude of this quantity is $\epsilon^{-\alpha}$, which is less than unity, and its phase angle is $-\beta$. If a vector V is multiplied by this complex quantity, the product $V\epsilon^{-\alpha-j\beta} = V\epsilon^{-\alpha}\epsilon^{-j\beta}$ has a magnitude $V\epsilon^{-\alpha}$, which is smaller than V , and has a phase angle $-\beta$ with respect to V . These relationships are indicated in Fig. 44.2, for β positive.

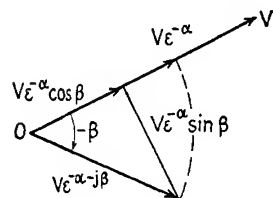


FIG. 44.2.—Relation between V and $V\epsilon^{-\alpha}\epsilon^{-j\beta}$.

45. Natural Logarithm of $(1+x)$.—Integration can sometimes be used to obtain the series development of a given function. For example, let

$$y = \log_e(1+x)$$

then

$$\frac{dy}{dx} = \frac{1}{1+x} = 1 - x + x^2 - x^3 + x^4 - \dots$$

Integrating term by term, using (34.3),

$$y = C + x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots \quad (45.1)$$

For $x = 0$, $\log_e(1+x) = \log_e 1 = 0$; hence, the integration constant C must be zero, and

$$\log_e(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots \quad (45.2)$$

This series is valid whenever the absolute value of x is less than unity.

When x is so small that $x \ll 1$ (x very small compared with unity), (45.2) gives the approximation formula

$$\log_e(1+x) \doteq x \quad (45.3)$$

VI. LINEAR DIFFERENTIAL EQUATIONS WITH CONSTANT COEFFICIENTS: FIRST ORDER

46. Definitions.—A relation between an *unknown* function y of a variable x and its derivatives with respect to x is called a *differential equation*. This relation may include known functions of

x , say $f(x)$, $\phi(x)$, The following four types of equations are sufficient for solving many circuit problems.

$$A \frac{dy}{dx} + By = 0 \quad (46.1)$$

$$A \frac{dy}{dx} + By = f(x) \quad (46.2)$$

$$A \frac{d^2y}{dx^2} + B \frac{dy}{dx} + Cy = 0 \quad (46.3)$$

$$A \frac{d^2y}{dx^2} + B \frac{dy}{dx} + Cy = f(x) \quad (46.4)$$

These are all *linear* because y , dy/dx , d^2y/dx^2 appear only in the first degree (like y and x in the ordinary equation of a straight line, $y = mx + p$). The phrase *with constant coefficients* is added because the coefficients A , B , C of y , y' , y'' on the left-hand side are all constants. On the right-hand side is $f(x)$, a known function of x , which may be identically zero but y must not appear. Equations (46.1) and (46.3) are said to be *without second member*, (46.2) and (46.4) *with second member*. Equations (46.1) and (46.2) are of the *first order*, because that is the order of the highest derivative present; (46.3) and (46.4) are of the *second order*.

The reason why linear differential equations with constant coefficients are the kind ordinarily encountered in circuit theory is that the final equations of an electric network, transmission line, antenna, however complicated, always arise from combinations of the four fundamental equations

$$e = L \frac{di}{dt} \quad e = Ri \quad q = Ce \quad i = \frac{dq}{dt}$$

relative to an inductance, a resistance, a capacitance. These equations are all linear, the coefficients being practically constant in many applications.

47. Superposition Principle.—A fundamental property of all linear differential equations with constant coefficients, of whatever order, is the superposition principle. Its *mathematical* meaning is the following:

Consider a linear differential equation with a second member, (46.4) for example,

$$A \frac{d^2y}{dx^2} + B \frac{dy}{dx} + Cy = f(x) \quad (47.1)$$

Consider two other differential equations with the same left-hand terms as (47.1), but one having a different, known function on the right-hand side,

$$A \frac{d^2y}{dx^2} + B \frac{dy}{dx} + Cy = \phi(x) \quad (47.2)$$

and the other having the *sum* of $f(x)$ and $\phi(x)$ on the right-hand side,

$$A \frac{d^2y}{dx^2} + B \frac{dy}{dx} + Cy = f(x) + \phi(x) \quad (47.3)$$

Let $y_1(x)$ be a solution of (47.1) and $y_2(x)$ a solution of (47.2). This means that

$$A \frac{d^2y_1}{dx^2} + B \frac{dy_1}{dx} + Cy_1 = f(x) \quad (47.4)$$

$$A \frac{d^2y_2}{dx^2} + B \frac{dy_2}{dx} + Cy_2 = \phi(x) \quad (47.5)$$

Consider now the function of x , $y_1(x) + y_2(x)$, and add the relations (47.4) and (47.5). Then

$$A \frac{d^2(y_1 + y_2)}{dx^2} + B \frac{d(y_1 + y_2)}{dx} + C(y_1 + y_2) = f(x) + \phi(x) \quad (47.6)$$

or the *sum* $y_1 + y_2$ is a solution of (47.3) of which the right-hand side is the *sum* $f(x) + \phi(x)$.

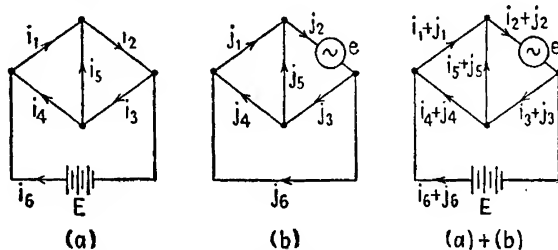


FIG. 47.1.—Illustrating the principle of superposition.

The *electrical* meaning of the superposition principle is that, in a circuit containing constant values of resistance, inductance, capacitance, the current caused by several emf's acting simultaneously is equal to the sum of the currents caused by each emf acting alone. For example, let each of the currents i_1, i_2, \dots, i_6 in the network of Fig. 47.1a be known when the constant emf E is applied in branch 6. Next, let each of the currents j_1, j_2, \dots, j_6 in the same net-

work, Fig. 47.1b, be known when an alternating emf e is inserted in branch 2. The superposition principle states that if both the emf's are applied simultaneously, or *superposed*, as in the figure marked $a + b$, the resulting current in branch 1 is $i_1 + j_1$, the current in branch 2 is $i_2 + j_2$, etc.

48. Example of First-order Equation.—The simplest example of a linear differential equation of the first order is

$$\frac{dy}{dx} = f(x) \quad (48.1)$$

an equation with second member. Its general solution is

$$y = F(x) + C \quad (48.2)$$

where $F(x)$ is any primitive function of $f(x)$ and C is the integration constant, Sec. 29.

49. Second Example of First-order Equation.—As the next example, consider

$$\frac{dy}{dx} - y = 0 \quad (49.1)$$

an equation without second member.

Suppose that, for $x = 0$, $y = C$. By (49.1), $dy/dx = C$ for $x = 0$. Taking the derivatives of both sides of (49.1),

$$\frac{d^2y}{dx^2} - \frac{dy}{dx} = 0 \quad (49.2)$$

whence the second derivative of y , and (for the same reason) all the derivatives of y , of all orders, are equal

to C for $x = 0$. Therefore, from Maclaurin's formula (41.1),

$$\begin{aligned} y &= C + Cx + C\frac{x^2}{2!} + C\frac{x^3}{3!} + \dots \\ &= C\left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots\right) \end{aligned} \quad (49.3)$$

The series in the brackets is the development (42.1) of e^x . Therefore,

$$y = Ce^x \quad (49.4)$$

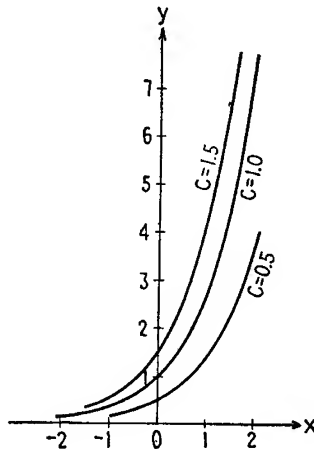


FIG. 49.1.—Graphs of three distinct integrals of equation (49.1).

Equation (49.4) represents an infinite number of functions, any one of which is obtained by giving C a definite numerical value. In the plane (Ox, Oy) these functions, the *integrals* of (49.1), are represented by an infinity of curves, each characterized by a definite value of C . C is the *integration constant*, or *parameter*. Figure 49.1 shows curves for the function $y = C\epsilon^x$ for three values of C .

Another way of defining a particular curve of the C family is to specify that, for $x = x_0$, $y = y_0$. (x_0, y_0) are called *initial conditions*. For example, if, for $x_0 = 1$, $y_0 = 1.359 = \frac{1}{2}(2.718)$, C must be 0.5 and $y = 0.5\epsilon^x$. Usually x_0 is taken equal to zero.

50. First-order Equation without Second Member.—The general type of the linear differential equation of the first order, with constant coefficients, without second member, is

$$A \frac{dy}{dx} + By = 0 \quad (50.1)$$

Assume a solution of the type $y = \epsilon^{px}$, where p is an unknown constant. Then

$$Ap\epsilon^{px} + B\epsilon^{px} = (Ap + B)\epsilon^{px} = 0$$

Since ϵ^{px} cannot be zero, Sec. 4,

$$Ap + B = 0 \quad \text{or} \quad p = -\frac{B}{A}$$

This is called the auxiliary equation.

The *general solution* of (50.1) is

$$y = C\epsilon^{-Bx/A} \quad (50.2)$$

which contains one arbitrary parameter C . The value of C is determined in each physical problem by the initial conditions (x_0, y_0) .

51. General First-order Equation with Second Member.—The general type of the linear differential equation of the first order, with second member, is

$$A \frac{dy}{dx} + By = \phi(x) \quad (51.1)$$

Its general solution is the sum of two functions:

1. One particular solution, $y_p(x)$, of (51.1).
2. The general solution of (51.1) *without* second member, that is, (50.1); this general solution is (50.2), and contains one arbitrary parameter C .

Thus the general solution of (51.1) is

$$y = y_p(x) + C\epsilon^{-Bx/A} \quad (51.2)$$

The first function $y_p(x)$ is called a *particular integral* (P.I.) of (51.1) and is, in electrical problems, the *steady-state solution* of (51.1); the second function is called the *complementary function* (C.F.) or the *transient solution*.

The terms "steady state" and "transient" are used because, in most network problems, A and B are positive constants and the variable x is time. The complementary function $C\epsilon^{-Bx/A}$ is equal to C for $x^1 = 0$ and diminishes to zero exponentially; it becomes negligibly small after a certain interval of time (say $x = 7A/B$, when the exponential becomes $\epsilon^{-7} \doteq 0.001$), and there remains only the steady-state solution $y_p(x)$.

Types of functions that are commonly encountered as second member $\phi(x)$ of (51.1) are a constant or a sinusoidal function of x (d-c or a-c voltage).

1. If $\phi(x) = m$, a constant, $y_p = m/B$, also a constant.

2. If $\phi(x) = \cos mx$ (or any function of the type $K \cos mx + L \sin mx$),

$$y_p(x) = M \cos mx + N \sin mx \quad (51.3)$$

where M and N are constants to be determined. Then $y_p(x)$ in (51.2) may be replaced by (51.3) and M and N obtained by equating the coefficients of $\cos mx$ and $\sin mx$ on both sides of the resulting equation.

VII. LINEAR DIFFERENTIAL EQUATIONS WITH CONSTANT COEFFICIENTS: SECOND ORDER

52. Second-order Equation without Second Member.—The second-order linear differential equation, with constant coefficients and without second member, is

$$A \frac{d^2y}{dx^2} + B \frac{dy}{dx} + Cy = 0 \quad (52.1)$$

Assume a solution of the type $y = \epsilon^{px}$, where p is an unknown constant; then

$$Ap^2\epsilon^{px} + Bp\epsilon^{px} + C\epsilon^{px} = (Ap^2 + Bp + C)\epsilon^{px} = 0$$

Since ϵ^{px} cannot be zero,

$$Ap^2 + Bp + C = 0 \quad (52.2)$$

Equation (52.2) is called the *auxiliary equation*. Three cases must be distinguished, according as the quadratic equation (52.2) has two real roots, two complex roots, one double (real) root.

53. Discussion.—Equation (52.2) has two roots,

$$k_1 = \frac{-B + \sqrt{B^2 - 4AC}}{2A} \quad k_2 = \frac{-B - \sqrt{B^2 - 4AC}}{2A} \quad (53.1)$$

1. If $B^2 - 4AC > 0$, both roots are real. The general solution of (52.1) is then

$$y = C_1 e^{k_1 x} + C_2 e^{k_2 x} \quad (53.2)$$

which contains *two* arbitrary parameters C_1, C_2 . The values of these parameters are determined in each physical problem by the *initial conditions*. In an equation of the second order, these are the values of the function y and of its first derivative dy/dx , for a given value of $x = x_0$ (usually $x_0 = 0$).

In case (1) the function $y(x)$ is *overdamped* and nonoscillatory.

2. If $B^2 - 4AC < 0$, the roots (53.1) are complex, conjugate (Sec. 20), and can be written

$$k_1 = -\alpha + j\omega \quad k_2 = -\alpha - j\omega \quad (53.3)$$

The general solution of (52.1) is then

$$y = C' e^{(-\alpha + j\omega)x} + C'' e^{(-\alpha - j\omega)x} \quad (53.4)$$

$$= e^{-\alpha x} (C' e^{j\omega x} + C'' e^{-j\omega x}) \quad (53.5)$$

where C', C'' are two arbitrary parameters. By Euler's formula (44.2),

$$e^{j\omega x} = \cos \omega x + j \sin \omega x \quad e^{-j\omega x} = \cos \omega x - j \sin \omega x \quad (53.6)$$

and (53.5) can be written

$$y = e^{-\alpha x} (C_1 \cos \omega x + C_2 \sin \omega x) \quad (53.7)$$

In case (2) the function $y(x)$ is *oscillatory*.

3. If $B^2 - 4AC = 0$ (critical case), both roots of (52.2) are equal to $-B/2A$, which is a real number. The general solution of (52.2) is then

$$y = (C_1 x + C_2) e^{-Bx/2A} \quad (53.8)$$

In every case the general solution of (52.1) contains two parameters, C_1, C_2 .

54. Second-order Equation with Second Member.—The general type of the linear differential equation of the second order, with second member, is

$$A \frac{d^2y}{dx^2} + B \frac{dy}{dx} + Cy = \phi(x) \quad (54.1)$$

As in the case of the first-order equation, Sec. 51, the general solution of (54.1) is the sum of two functions:

1. One particular integral, $y_p(x)$, of (54.1).
2. The complementary function, which is the general solution of (54.1) *without* second member,

$$y_c(x) = C_1 e^{k_1 x} + C_2 e^{k_2 x} \quad (54.2)$$

In practice, $\phi(x)$ is either a constant or a sinusoidal function of x (d-c or a-c voltage).

If $\phi(x) = m$, a constant, $y_p(x) = m/C$, also a constant. If $\phi(x) = \cos mx$, or any function of the type $K \cos mx + L \sin mx$, then, as in Sec. 51,

$$y_p(x) = M \cos mx + N \sin mx \quad (54.3)$$

and M and N are obtained by equating the coefficients of $\cos mx$ and $\sin mx$ on both sides of the equation.

If the symbolic notation of sinusoidal quantities is used, this last solution takes a slightly different form. Suppose $\phi(x) = \hat{E} \cos mx$, which is the real part of $\hat{E} e^{jmx}$. Consider the differential equation

$$A \frac{d^2y}{dx^2} + B \frac{dy}{dx} + Cy = \hat{E} e^{jmx} \quad (54.4)$$

and assume a solution of the type $Q e^{jmx}$. Then

$$(-Am^2 + Bj m + C)Q = \hat{E}$$

and

$$Q = \frac{\hat{E}}{(C - Am^2) + jBm} \quad (54.5)$$

In symbolic notation $Q e^{jmx}$ is the particular integral or steady-state solution sought. The actual solution, alternating function of time, is the *real* part of $Q e^{jmx}$, or

$$\frac{C - Am^2}{(C - Am^2)^2 + B^2 m^2} \hat{E} \cos mx \quad (54.6)$$

Either method is advantageous depending upon the problem at hand.

55. Circular Functions.—Consider the equation

$$\frac{d^2y}{dx^2} + y = 0 \quad (55.1)$$

of which the auxiliary equation is $p^2 + 1 = 0$. Its general solution is

$$y = C'e^{ix} + C''\epsilon^{-ix} \quad (55.2)$$

Let $y_1(x)$ be the solution of (55.1) defined by the initial conditions $y_1(0) = 1$, $y_1'(0) = 0$. Then,

$$\left. \begin{aligned} C' + C'' &= 1 \\ jC' - jC'' &= 0 \end{aligned} \right] ;$$

hence

$$y_1(x) = \frac{e^{ix} + \epsilon^{-ix}}{2} \quad (55.3)$$

From Euler's formulas (53.6), for $\omega = 1$,

$$\cos x = \frac{e^{ix} + \epsilon^{-ix}}{2} \quad \sin x = \frac{e^{ix} - \epsilon^{-ix}}{2j} \quad (55.4)$$

Therefore the solution y_1 of (55.1) is $y_1 = \cos x$.

In the same way $\sin x$ can be defined as that solution $y_2(x)$ of (55.1) which satisfies the initial conditions $y_2(0) = 0$, $y_2'(0) = 1$.

The general solution of (55.1) then can be written, instead of (55.2), as

$$y = C_1 \cos x + C_2 \sin x \quad (55.5)$$

where C_1 , C_2 are arbitrary parameters.

56. Hyperbolic Functions.—Consider the differential equation

$$\frac{d^2y}{dx^2} - y = 0 \quad (56.1)$$

of which the auxiliary equation is $p^2 - 1 = 0$. Its general solution is

$$y = C'\epsilon^x + C''\epsilon^{-x} \quad (56.2)$$

Let $y_1(x)$ be the solution of (56.1) defined by the initial conditions $y_1(0) = 1$, $y_1'(0) = 0$; let $y_2(x)$ be defined by $y_2(0) = 0$, $y_2'(0) = 1$. Then y_1 and y_2 are the *hyperbolic cosine* and the *hyperbolic sine* of x ,

and

$$\cosh x = \frac{e^x + e^{-x}}{2} = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots \quad (56.3)$$

$$\sinh x = \frac{e^x - e^{-x}}{2} = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots \quad (56.4)$$

These developments are valid for all values of x , real or complex.

The graphs of these functions are shown in Fig. 56.1 and in Fig. 56.2. They are not periodic. However, from (53.6),

$$e^{2j\pi} = 1 \quad (56.5)$$

so that adding $2j\pi$ to x changes neither $\cosh x$ nor $\sinh x$. These functions therefore admit the imaginary period $2j\pi$.

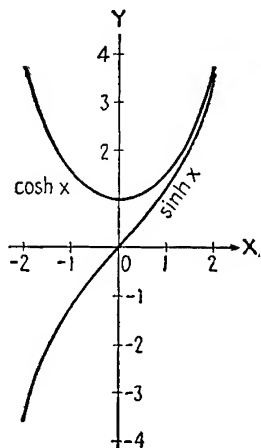


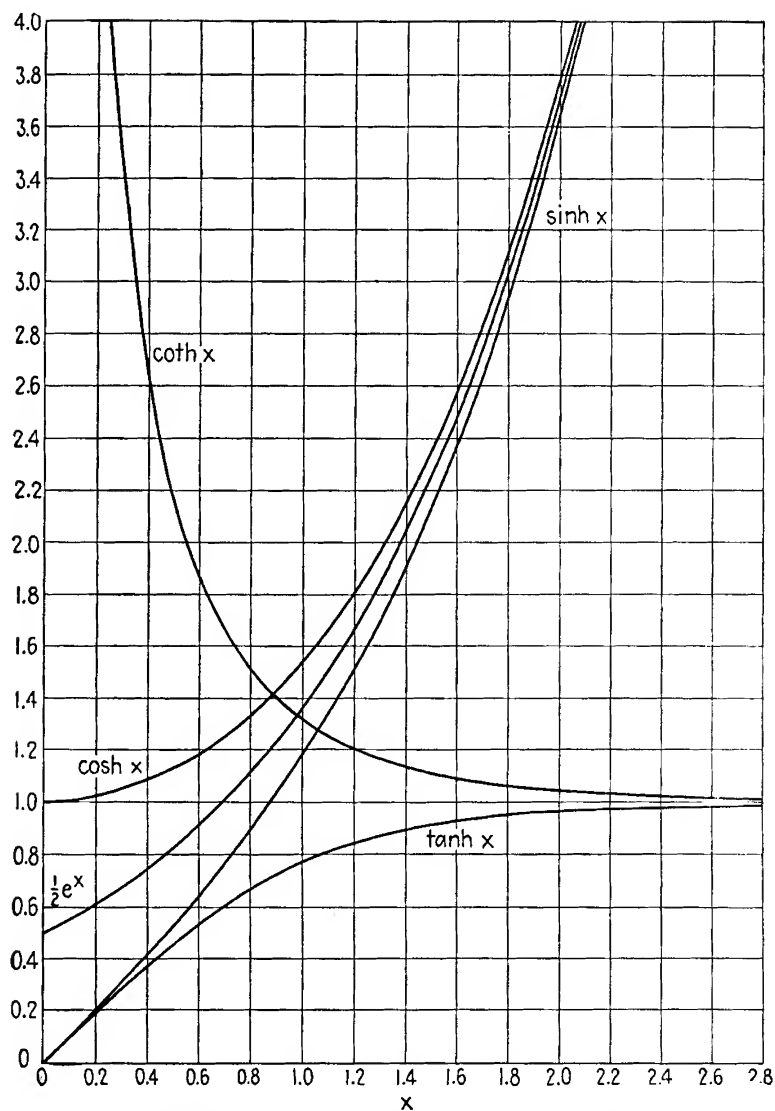
FIG. 56.1.—Graph of $\sinh x$ and $\cosh x$.

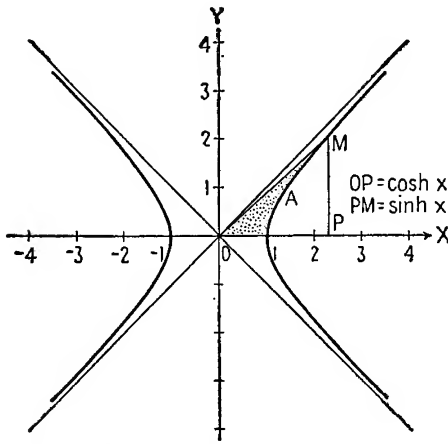
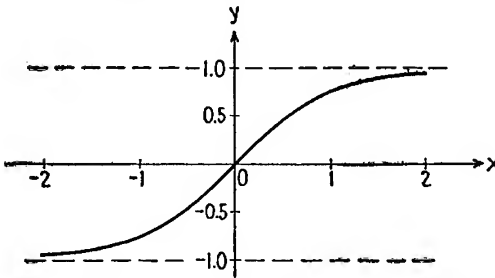
From (56.3) and (56.4),

$$(\cosh x)^2 - (\sinh x)^2 = 1 \quad (56.6)$$

Therefore, if $OP = \cosh x$, $PM = \sinh x$, the locus of M is a hyperbola, Fig. 56.3, whence the names.

The shaded area OAM can be shown to be equal to $x/2$, just as a sector of angle α radians in a circle of radius unity has for area $\alpha/2$. The pure number x is sometimes said to represent x *hyperbolic radians*; it should not be mistaken for the geometrical angle MOA . If x is doubled, the shaded area is doubled, not the angle. Note that hyperbolic radians and nepers are pure numbers.

FIG. 56.2.—Hyperbolic functions of a real variable x .

FIG. 56.3.—Rectangular hyperbola as locus of $\sinh x$ and $\cosh x$.FIG. 56.4.—Graph of the function $y = \tanh x$.

The *hyperbolic tangent*, Fig. 56.4, is

$$\begin{aligned} \tanh x &= \frac{\sinh x}{\cosh x} \\ &= \frac{e^x - e^{-x}}{e^x + e^{-x}} \end{aligned} \quad (56.7)$$

57. Circular and Hyperbolic Functions of Complex Quantities.—

The correspondence between circular and hyperbolic functions is obtained from (55.4), (56.3), (56.4).

$$\left. \begin{aligned} \sin jx &= j \sinh x \\ \cos jx &= \cosh x \end{aligned} \right] \quad (57.1) \quad \left. \begin{aligned} \sinh jx &= j \sin x \\ \cosh jx &= \cos x \end{aligned} \right] \quad (57.2)$$

Addition formulas for the hyperbolic functions are

$$\left. \begin{aligned} \sinh (a \pm b) &= \sinh a \cosh b \pm \sinh b \cosh a \\ \cosh (a \pm b) &= \cosh a \cosh b \pm \sinh a \sinh b \end{aligned} \right] \quad (57.3)$$

With the help of these formulas the real and imaginary parts of the \sinh or \cosh of a complex quantity can be obtained. Upon writing α instead of a and $j\beta$ instead of b and using (57.2), formulas (57.3) become

$$\left. \begin{aligned} \sinh(\alpha + j\beta) &= \sinh \alpha \cos \beta + j \cosh \alpha \sin \beta \\ \cosh(\alpha + j\beta) &= \cosh \alpha \cos \beta + j \sinh \alpha \sin \beta \end{aligned} \right] \quad (57.4)$$

which are of constant use in calculations on uniform transmission lines. α is in hyperbolic radians or nepers, but not decibels; β is in circular radians.

Example.—To find the \sinh and \cosh of

$$\alpha + j\beta = 2.72 - j0.474$$

Tables give

$$\sinh 2.72 = 7.55722 \quad \cosh 2.72 = 7.62310$$

Next $0.474/\pi = \phi^\circ/180$; hence $\phi = 27^\circ 9.5'$, and, from tables,

$$\sin 27^\circ 9.5' = 0.45645 \quad \cos 27^\circ 9.5' = 0.88975$$

and this angle is to be taken negatively. Hence, from (57.4), keeping four significant figures only,

$$\begin{aligned} \sinh(2.72 - j0.474) &= 7.557 \cdot 0.8897 - j7.623 \cdot 0.4565 \\ &= 6.723 - j3.479 \\ \cosh(2.72 - j0.474) &= 7.263 \cdot 0.8897 - j7.557 \cdot 0.4565 \\ &= 6.782 - j2.543 \end{aligned}$$

58. Inverse Circular and Hyperbolic Functions.—If a complex quantity $\alpha + j\beta$ is given that is the \sin , or \cos , or \sinh , or \cosh of an unknown quantity γ , it is possible to calculate the real and imaginary parts of $\gamma = \alpha + j\beta$. As an example, the solution of the equation $\cosh \gamma = v$ will be given, where v is a real number. This equation appears in the theory of filters and lines. There are three cases.

1. $v < -1$. Putting $\beta = \pi$ in the second equation (57.4),

$$\cosh(\alpha + j\pi) = -\cosh \alpha \quad (58.1)$$

Therefore $\gamma = \alpha + j\pi$; and, from Fig. 56.1, as v varies from $-\infty$ to -1 , $\cosh \alpha$ varies from $+\infty$ to 1 , and α varies from $+\infty$ to zero, while β remains constant and equal to π .

2. $-1 < v < 1$. Putting $\alpha = 0$ in the second equation (57.4),

$$\cosh(0 + j\beta) = \cos \beta \quad (58.2)$$

yielding the second equation (57.2). Therefore $\gamma = j\beta$; and as

v or $\cos \beta$ varies from -1 to $+1$, β varies from $+\pi$ to zero, while α remains equal to zero.

3. $1 < v$. γ is then a real number α ; and as v or $\cosh \alpha$ varies from 1 to $+\infty$, α varies from 0 to $+\infty$, while β remains equal to zero.

To sum up this discussion, when $\cosh \gamma$ goes from $-\infty$ to $+\infty$, taking all real values, $\gamma = \alpha + j\beta$ describes in the complex plane the path indicated in Fig. 58.1.

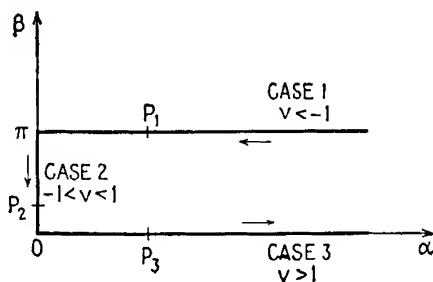


FIG. 58.1.—Graph of the solutions of the equation $\cosh x = v$.

For example, at P_1 ,

$$\cosh \gamma = -12.29 \quad \gamma = 3.20 + j\pi$$

At P_2 ,

$$\cosh \gamma = 0.500 \quad \gamma = j\frac{\pi}{3} = j1.04$$

At P_3 ,

$$\cosh \gamma = 12.29 \quad \gamma = 3.20$$

59. Derivatives of Hyperbolic Functions.—The following formulas are given here to supplement those of Sec. 24:

Function	Derivative	
$\sinh x$	$\cosh x$	(59.1)
$\cosh x$	$\sinh x$	(59.2)
$\sinh^{-1} x$	$\frac{1}{\sqrt{1+x^2}}$	(59.3)
$\cosh^{-1} x$	$\frac{1}{\sqrt{x^2-1}}$	(59.4)

After two differentiations of $A \sinh x + B \cosh x$, the original function returns (cf. Sec. 28).

APPENDIX B

BRIEF REVIEW OF ELECTRICITY AND MAGNETISM

The purpose of this appendix is to review certain basic concepts concerning the motion of electrons in wires and in evacuated spaces. The principles here discussed apply to circuits whose linear dimensions are small enough so that the action of one part of a circuit on any other part may be considered instantaneous. When the linear dimensions are large, there is an appreciable time delay between cause and effect, the effect of the motion of an electron being felt at a distant point some time after the electron has moved, the delay depending upon the distance. This is of particular significance in antennas and wave guides.¹

1. Mks System of Units.—In an absolute system of units, a number of basic independent units are chosen arbitrarily as fundamental and all other units are derived from them. In the system of units used in this book the fundamental units are units of length, mass, time, electric charge. In order that the mechanical and electrical units shall have convenient magnitudes, the fundamental unit of length chosen is the meter, of mass the kilogram, of time the second, of charge the coulomb (Sec. 2). This system is known as the meter-kilogram-second (mks) system. [As yet, the choice of the coulomb as a fourth unit has not been approved by the International Electrotechnical Commission (I.E.C.).] A special form of this system, the “rationalized” mks system, is used here.²

The four fundamental units and a few of the simpler derived units are listed in Table 1.1, and other units will be defined as needed.

It will be shown that simple relationships exist between the fundamental units and the practical electrical units, ampere, volt, ohm, henry, farad.

¹ R. W. P. KING, H. R. MIMNO, and A. H. WING, “Transmission Lines, Antennas, and Wave Guides,” McGraw-Hill Book Company, Inc., 1945.

² In the rationalized system, the proportionality factor 4π is absent from most of the equations, but appears in the equations defining forces between charges and between current elements.

TABLE 1.1

Quantity	Symbol	Mks unit	Explanatory comparison
Length	L	Meter	39.37 in.; slightly more than 1 yd
Mass	m	Kilogram	1,000 grams, approximately 2.2 lb-mass
Time	t	Second	
Electric charge	q	Coulomb	Defined in Sec. 2
Force	F	Newton	10^5 dynes; approximately 3.5 oz
Velocity	v	Meter/sec	1 meter/sec equals 2.23 miles/hr
Energy or work	W	Joule	10^7 ergs; approximately $\frac{1}{4}$ ft-lb
Power	P	Watt	1 joule/sec; 746 watts equals approximately 1 hp

2. Electron; Electric Charge.—The concept of electric charge is based upon experiment. It involves particles that exert forces on each other depending upon the distance between them. Experiments indicate that matter is made up of two kinds of electric charge, designated as positive and negative, each atom of matter normally having equal amounts of positive and negative charge. The exterior parts of the atom contain only negative charges, existing in discrete units called electrons. All electrons are identical.

Because all electrons have the same electric charge and subdivision of this quantity never has been observed, the charge on an electron is a natural unit although often far too small for practical use. The “practical” unit of charge, the *coulomb*, is equal to the combined charge of $6.242 \cdot 10^{18}$ electrons; electronic charge $e = 1.602 \cdot 10^{-19}$ coulomb. When accelerated at low speeds, the electron behaves as if it had a mass $m = 9.107 \cdot 10^{-31}$ kg.

Positive charge occurs always in multiples of the amount of the electronic charge. An object that is electrically neutral has equal quantities of the two kinds of charge. By the expenditure of a certain amount of energy it is possible to remove one or more electrons from an atom or molecule. The remaining part of the molecule is called a positive ion. Ions contain practically all the mass of the atom from which they are derived. The term “ion” is also applied generally to any small particle that acquires a negative or a positive charge by the addition or subtraction of electrons.

The removal of electrons from atoms or molecules may be accomplished by thermal energy (thermionic emission), by impact (ionization in a gas tube and secondary emission from a surface), by radiant energy (photoelectric emission from a surface and

ionization in the upper atmosphere). There are other means but these are the more important ones in electronics and radio.

At high temperatures, the electrons of the atoms constituting a piece of material may be agitated so violently that some of them are ejected through the surface of the material. This process is known as *thermionic emission*. Electrons thus freed form the current through the evacuated space in the thermionic vacuum tube, Chap. X.

Air molecules in the extremely rarefied upper atmosphere may be split by sunlight into free electrons and positive ions, which may wander about for seconds, minutes, hours (depending upon the rarity of the atmosphere) before recombination occurs. As a result there are strata in the upper atmosphere that contain free electrons and ions. The height and the degree of ionization of these strata vary in various cycles of time. These regions in the *ionosphere* refract and reflect radio signals.¹

Electric charges can be neither created nor destroyed. In any closed electrical system no process changes the total net amount of electric charge present. This is known as the *principle of the conservation of electric charge*.

3. Coulomb's Law, Electric Field and Potential.—The following experiment is fundamental. Let A represent a fixed and very small body from which some electrons have been removed so that A has a net positive charge. A similar body B , also positively charged, is repelled when brought near A . The force of repulsion acts along a line joining the two bodies and is given by *Coulomb's law*

$$f = K \frac{Q_A Q_B}{d^2} \quad (3.1)$$

where Q_A and Q_B are the charges on the very small bodies A and B ; d is their distance apart, which must be large compared with the dimensions of the small bodies; K is a factor of proportionality depending upon the units used. If Q_A and Q_B are in coulombs, d in meters, f in newtons, K is equal to $1/4\pi\epsilon_0$ in vacuum or air and to $1/4\pi\epsilon$ in a homogeneous dielectric medium, where ϵ_0 and ϵ are constants defined in Sec. 5.

A force of mutual repulsion is likewise observed when both bodies are negatively charged. If the bodies are oppositely charged, the force (3.1) is one of attraction. In brief, like charges repel, unlike charges attract each other.

¹ KING, MIMNO, and WING, *op. cit.*, Chap. IV.

Direct application of Coulomb's law when the number of charges is very large is too cumbersome for practical use. Considerable simplification is often achieved by introducing the concept of *electric field*. For the static condition, (3.1) in mks units is

$$F = \frac{1}{4\pi\epsilon} \frac{Q_A Q_B}{d^2}$$

which may be arranged as

$$F = Q_A \left(\frac{1}{4\pi\epsilon} \frac{Q_B}{d^2} \right)$$

or

$$\mathbf{F} = Q_A \mathbf{E} \quad (3.2)$$

where \mathbf{E} is the force per unit charge in magnitude and direction exerted on Q_A by the charge Q_B . This concept may be extended to include other stationary or moving charges. Thus, an electric field is said to exist in a region of space where a charge experiences a force that depends upon the magnitude and position of the charge. The direction of the field at a point is the direction of the force on a positive charge at that point or opposite to the force on a negative charge. The magnitude of the field at a point, or "field intensity," is the magnitude of the force per unit charge. At a point where there is no charge the electric field intensity is the force per unit charge that would be exerted on a positive charge placed at that point if the introduction of the charge would not significantly alter the position or motion of other charges.

If a small charged body is moved slowly in a steady electric field (all other charges remaining unchanged in magnitude and position), work may be done by or against electric forces. If the work done against the electric forces in moving a positive charge from point a to point b is one joule per coulomb, points a and b are said to differ in potential by one volt, point b being one volt higher in potential than point a . Difference of potential is denoted¹ by E , work by W , and

$$W = QE \quad (3.3)$$

If a small charged body is moved along a path that is everywhere perpendicular to \mathbf{E} , no work is performed on the charge and the charge is said to have moved on an equipotential surface. The surface of a perfect conductor is an equipotential surface.

An electric field is represented graphically by lines of force

¹ The symbol V is also used for difference of potential.

whose direction at every point is the direction of the field at that point and whose number per unit area perpendicular to these lines is proportional to the average strength of the field over the area. Thus, in regions where the field strength is great, the lines are close together. Lines of force are perpendicular (normal) to equipotential surfaces.

The dashed lines of Fig. 3.1 represent a section across three equipotential surfaces of a steady electric field. The solid line $P'PP''$ represents a line of force. The potential of the surface containing P is taken as the reference for potential and is marked 0 (zero). The surface containing point P' is $|E|$ volts higher in potential than P and is marked $+|E|$. The surface containing point P'' is $|E|$ volts lower in potential than P and is marked $-|E|$.

The electric field \mathcal{E} is directed toward the surface of lower potential and away from the surface of higher potential since work is done

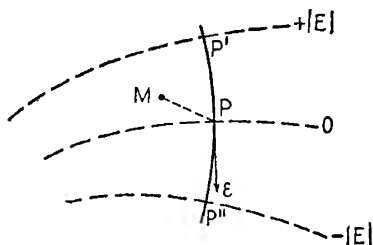


FIG. 3.1.—Indicating surfaces of constant potential and direction of electric field \mathcal{E} .

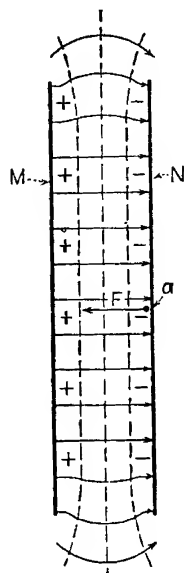


FIG. 3.2.—Cross section of plane parallel conducting surfaces in vacuum.

against the electric force when a positive charge is moved to a point of higher potential. At P the work done per unit distance is greater along the line of force PP' than along any other line such as PM because motion normal to equipotential surfaces is collinear with the force due to the electric field. The *potential gradient* at a point such as P is the space rate of change of potential in the direction of the greatest rate of increase. Thus the potential gradient at P is normal to the equipotential surface at P and is directed toward the surface of higher potential. The electric field \mathcal{E} at P is also normal to the equipotential surface at P and is directed toward the

surface of lower potential. Thus potential gradient and electric field are equal in magnitude but are oppositely directed. Field strength in newtons/coulomb and magnitude of potential gradient in volts/meter are numerically equal.

A central cross section of two plane parallel conducting surfaces M and N in a vacuum is shown in Fig. 3.2. When electrons are removed from M and transferred to N , a potential difference E and an electric field are established between M and N . If the plate separation d is small compared with the surface dimensions of the plates, the lines of force in the middle of the plate area are straight and parallel. In this region the field is uniform and has the same strength, $\mathcal{E} = -E/d$, at every point. Cross sections of several equipotential surfaces are indicated by dashed lines, Fig. 3.2, these surfaces being perpendicular to the lines of force at every point.

4. Free Electron in Electric Field.—If an electron were released from point a , Fig. 3.2, on plate N near its center, the electron would be accelerated toward M and would move along a straight line perpendicular to M and N owing to a constant force

$$\mathbf{F} = -e\mathcal{E} \text{ newtons} \quad (4.1)$$

The plates being in vacuum, there would be no opposition to the movement of the electron and all the work eE done on it by the field would be converted into kinetic energy. If the electron were released without initial velocity, the final kinetic energy, immediately before impact against plate M , would be

$$\frac{1}{2}mv^2 = eE \text{ joules} \quad (4.2)$$

where m is the mass of the electron in kilograms, v is its final velocity in meters/sec, E is the potential difference between the plates in volts. Rearranging terms,

$$v = \sqrt{2 \frac{e}{m} E} \quad (4.3)$$

The mass m of the electron increases with velocity. However, for potential differences less than 5,000 volts, m differs by less than 1 per cent from the value given in Sec. 2, and the velocity v is

$$v \doteq 5.93 \cdot 10^5 \sqrt{E} \text{ meters/sec} \quad (4.4)$$

The distance between the plates is not a factor in determining the final velocity. Since all electrons have the same charge and mass, they acquire the same energy in a given field. This energy depends

solely upon the potential difference through which the electrons move.

In passing through a potential difference of 1 volt an electron acquires an energy of $1.60 \cdot 10^{-19}$ joule. This is a convenient unit of energy in electronics and is called the *electron volt*.

In Fig. 4.1 is shown a cross section of a uniform electric field in vacuum with the lines of force lying in the plane of the paper and directed as shown.¹ When a moving electron enters such a field perpendicularly, it experiences a force of magnitude $F = e\mathcal{E}$, in a direction opposite to the field. As the force F remains constant in magnitude and direction, the electron follows a parabolic path within the field. The electron leaves the field with a slightly increased velocity v , the vector sum of the original velocity v_x and the acquired velocity v_y parallel to the field. After leaving the field

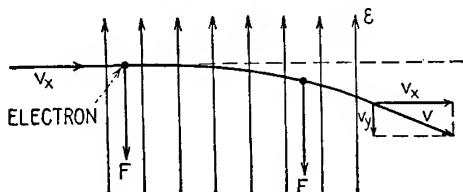


FIG. 4.1.—Deflection of an electron in a uniform electric field.

the electron travels in a straight line in the direction of emergence with the constant velocity v .

5. Capacitance and Capacitor.—The potential difference E established between the plates, Fig. 3.2, is directly proportional to the quantity of charge Q transferred from one plate to the other. The plates constitute a simple two-plate *capacitor*. For any capacitor of this type the ratio of Q to E is a constant, and by definition this constant is its *capacitance* C .

$$C = \frac{Q}{E} \quad (5.1)$$

When Q is in coulombs and E in volts, C is in *farads*. The transfer of one coulomb of charge would produce a potential difference of one volt between the terminals of a capacitor of one farad capacitance. A capacitance of one farad is very much greater than is encountered in practice; actual capacitances are measured in microfarads (μf) or in micromicrofarads ($\mu\mu f$).

¹ In practice, the field would not begin and end abruptly as shown. The deflecting force at the edges would be smaller, but the effect would be nearly the same.

A capacitor consisting of two parallel plates of area A in square meters separated by a distance d in meters in vacuum has a capacitance approximately

$$C \doteq \epsilon_0 \frac{A}{d} \text{ farads} \quad (5.2)$$

provided that d is very small compared with the dimensions of the plates. Here, ϵ_0 is a universal electric constant, the dielectric constant, or *permittivity*, of free space, and its value is

$$\epsilon_0 \doteq 8.85 \mu\text{mf}/\text{meter} \quad (5.3)$$

If the space between the plates is filled with a dielectric, such as paraffined paper, porcelain, glass, mica, oil, a greater quantity of charge must be transferred from one plate to the other to establish a given potential difference. This is because an internal electronic rearrangement occurs within the dielectric that tends to neutralize partly the charges built up on each plate. When the dielectric is present, the capacitance is ϵ_r times the magnitude of the capacitance when there is vacuum between the plates, or

$$C \doteq 8.85\epsilon_r \frac{A}{d} \mu\text{mf} \quad (5.4)$$

The constant ϵ_r is the relative dielectric constant, a pure numeric, commonly called the dielectric constant of the dielectric material employed. Its value ranges from 2 to 10 for common dielectric materials and is about 80 for water. The specific inductive capacity, absolute dielectric constant, or permittivity of the material is $\epsilon = \epsilon_0\epsilon_r$ and is expressed in farads/meter.

The energy stored in a charged capacitor is

$$\text{Energy in capacitor} = \frac{1}{2}CE^2 \text{ joules} \quad (5.5)$$

where C is in farads and E in volts. Thus, half a joule is stored in a 1- μf capacitor when the potential difference is 1,000 volts.

When capacitors are employed in parallel, their capacitances are numerically additive. When employed in series, the reciprocal of the series capacitance is equal to the sum of the reciprocals of the individual capacitances.

$$\frac{1}{C_{\text{series}}} = \frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \cdots \quad (5.6)$$

6. Electric Current.—When electric charges move, an electric current exists. Experiments show that the effect of the motion of a

positive charge in one direction is the same as the effect of an equal motion of an equal negative charge in the opposite direction. Before the electronic nature of electricity was understood, all the electric and electromagnetic conventions were set up on the basis of positive charge and its movement. Therefore, the effects of an electric current are described in terms of a *conventional current* due to the motion of positive charge. By this convention, the current produced by an applied electric field or voltage is from a point of higher potential to a point of lower potential or from a battery terminal conventionally marked $+$ through an external circuit to the terminal conventionally marked $-$.

In metal wires, the current is due exclusively to the motion of the electrons, since the atoms and ions remain practically fixed owing to the solid nature of the metal; the average velocity causing the electric current is small, of the order of cm/sec. The current between the electrodes in a vacuum tube, if all the residual gas is removed, also is due exclusively to the motion of electrons. In ionized gases, the motion of the positive ions is relatively slow because the ions are heavy, whereas the motion of the electrons is relatively rapid since their mass is small. Hence most of the effects of current in an ionized gas are due to the motion of electrons.

When charged particles flow through a given surface or cross section at a uniform rate and transport one coulomb of charge per second through the surface, the magnitude of the current through the surface is one *ampere*. If the rate of flow of electric charge varies with time, a more general definition of the instantaneous current is

$$i = \frac{dq}{dt} \text{ amperes} \quad (6.1)$$

where q is in coulombs and t in seconds.

When positive charge moves from one point to another point at a higher potential, the rate at which work is done, (3.3), against the electrical forces is $p = ei$ watts, where p is the instantaneous power in watts, e is the instantaneous value of the potential difference in volts, and i is the instantaneous value of the current in amperes.

Small vacuum tubes frequently carry plate currents of the order of 1 ma. One milliampere is a flow of approximately $6 \cdot 10^{15}$ electrons/sec. The rate of arrival of the electrons at the plate is not uniform. . . . Under certain conditions, the fluctuation in the plate current may be translated into sound, called *fluctuation noise* or hiss.

The term fluctuation noise has been extended to include other similar disturbances and to include other than audible frequencies. These fluctuations limit the amplification that may be usefully employed. In the more extreme cases the impact of individual electrons on the plate can be detected.

7. Electric Resistance.—The electric resistance of a conductor is a measure of the difficulty that electrons experience as they are forced through the conductor under the influence of an electric field within the conductor. For most conductors, the ratio of E , the potential difference between the conductor terminals, to the corresponding steady current I is practically a constant (Ohm's law). The ratio between the potential difference E across a conductor in volts and the current I through it in amperes defines its resistance R in ohms and its conductance G in mhos.

$$R = \frac{E}{I} \text{ ohms} \quad \text{and} \quad G = \frac{I}{E} \text{ mhos} \quad (7.1)$$

These relations may be written

$$E = RI \quad \text{and} \quad I = GE \quad (7.2)$$

The resistance of a conductor is the voltage across it per ampere of current through it, while its conductance is the current through it per volt of potential applied.

The power in watts delivered to such a resistive conductor is

$$P = EI = I^2R = E^2G \text{ watts} \quad (7.3)$$

A wire of length l in meters and of cross section A in square meters has a resistance R in ohms given by

$$R = \rho \frac{l}{A} \quad (7.4)$$

where ρ is the *resistivity* of the material of the wire. A normal value of ρ for soft-drawn copper is $1.72 \cdot 10^{-8}$ ohm meter.

When conductors are in series, the total resistance is the sum of the individual resistances. When arranged in parallel the total resistance is given by the relation

$$\frac{1}{R_{\text{total}}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \cdots \quad (7.5)$$

8. Electromotive Force; Electric Circuit.—The open-circuit potential difference between the terminals of a source of electrical

energy is its *electromotive force* (emf), by definition. The current that a source of electric energy produces in a circuit is given by

$$I = \frac{\text{emf of energy source}}{\text{series resistance of entire circuit}}^1 \quad (8.1)$$

The various points in an electric circuit may be identified in terms of the potential at these points with respect to the potential of some reference point, such as the ground, a metal chassis, or an airplane fuselage. The reference potential is commonly referred to as *ground potential*, although the potential of the reference point with respect to the earth may be considerable.

Kirchhoff's laws, which are useful in the analysis of complicated circuits, may be stated as follows:

1. The algebraic sum of the currents that meet at any point is zero.
2. The algebraic sum of the potential differences (taken in the same direction) around any closed circuit is zero.

Equation (8.1) is Kirchhoff's second law applied to a simple circuit.

9. Electromagnetism.—It is found experimentally that two parallel and electrostatically neutral wires attract each other when carrying steady currents in the same direction. If the currents are in opposite directions, the wires repel each other. In general, current-carrying loops are subject to mechanical forces that would not exist if no currents were flowing. Forces, such as these, that depend upon the magnitude and direction of electric currents are called magnetic forces, and currents experiencing such forces are said to be in a *magnetic field* or *field of magnetic induction*. The field of magnetic induction is denoted by the vector B , defined as follows:

If a straight current element is used as a probe and is so oriented in a uniform magnetic field that the force on the current is maximum, the current, B field, and the force are mutually perpendicular and are oriented as the middle finger, index finger, and thumb of the left hand when these are held mutually perpendicular. This is known as Fleming's left-hand rule, Fig. 9.1. The magnitude of the force is proportional to the length l of the probe and to the current i that it carries. Writing

$$F \text{ (newtons)} = B \cdot l \text{ (meters)} \cdot i \text{ (amperes)} \quad (9.1)$$

the coefficient of proportionality B defines the field of magnetic

¹ Including the internal resistance of the source, Appendix C.

induction at the point where the probe is located. The magnitude of B thus is measured in newtons/meter-ampere.

In air, the direction of the vector B is the direction in which the north-seeking pole of a very small magnetic compass points, if free to rotate in all directions.

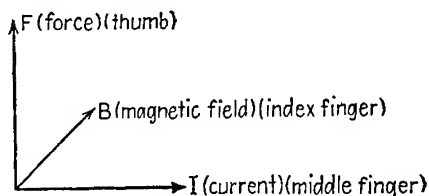


FIG. 9.1.—Illustrating Fleming's left-hand rule for force on a current element. The magnetic field is directed away from the observer.

10. Magnetic Flux.—A very small loop of conductor having area A is shown in Fig. 10.1. After a positive direction along the conductor has been specified, as shown, the direction of the positive normal to the small area is defined by the direction in which a right-handed screw would advance if turned in the positive direction. Call B_n the component of the magnetic induction B in the direction of the positive normal. The product of B_n by the area of the loop A ,

$$\phi = B_n A \quad (10.1)$$

is called the *magnetic flux* ϕ of B through the loop. The unit of flux is called the *weber*.

The direction and strength of the induction field B may be pictured by "lines of flux" whose direction represents the direction of the field and whose number per unit area perpendicular to the lines is proportional to the average magnitude of B over the area.

11. Magnetic Intensity.—Figure 11.1 represents an axial cross section of a long uniform solenoid in air. Lines of magnetic flux representing the field of magnetic induction are shown threading the coil. Toward the middle of the coil the magnetic flux is parallel to the axis of the coil and has uniform density. B is uniform in this region, and its magnitude is

$$B = \mu_0 \cdot i \text{ (amperes)} \cdot n_1 \text{ (turns/meter)} \quad (11.1)$$

where μ_0 is a universal constant, the *absolute permeability* of free

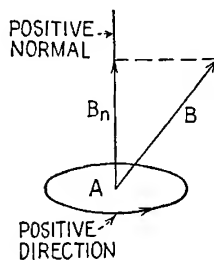


FIG. 10.1.—Conventions for determining flux of induction ϕ .

space. The product $i \cdot n_1$ (ampere turns/meter) is a measure of a physical quantity called *magnetizing force* or *magnetic intensity*, denoted by H .

The magnetic induction, or B field, and the magnetizing force, or H field, in vacuum have the same direction, and their numerical magnitudes have the same ratio at every point. This ratio is

$$\begin{aligned} \mu_0 = \frac{B}{H} \Big|_{\text{vacuum}} &\doteq 4\pi 10^{-7} \text{ newtons/ampere}^2 \text{ or henrys/meter} \\ &\doteq 1.26 \mu\text{h/meter} \end{aligned} \quad (11.2)$$

The constant μ_0 is *not* dimensionless, since the B and H units do not have the same dimensions.

When the medium is a “magnetic” substance (such as iron), the actual flux density B_m may be many times greater than B_a , the value it would have if the medium were air or vacuum. Assuming proportionality of B_m to B_a , which is approximately true when the magnetic medium is far



FIG. 11.1.—Magnetic field within a solenoid.

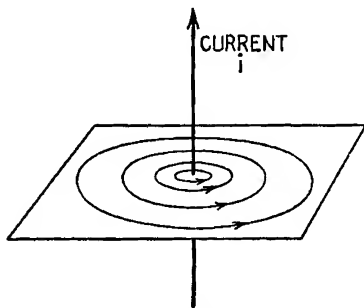


FIG. 11.2.—Magnetic field about a long straight current.

from magnetic saturation, the ratio B_m/B_a is a pure numeric, characteristic of the substance, called its *relative permeability*, or simply *permeability*, and is denoted by μ_r .

$$\mu_r = \frac{B_m}{B_a} \quad (11.3)$$

The value of μ_r ranges from a few hundred for commercial iron up to several hundred thousand for special magnetic alloys. If a coil is wound on a toroidal iron core, the magnetic flux density B_m within the iron is μ_r times the value it would have if the iron were absent.

The *absolute permeability* μ of a magnetic material is $\mu_0\mu_r$ henrys/meter.

The magnetic field about a long straight current is everywhere at right angles to the current. The lines of force are circles concentric with the current, Fig. 11.2. The magnetizing force at

distance r from the center of the current, outside of the current, is

$$H = \frac{i}{2\pi r} \text{ ampere turns/meter} \quad (11.4)$$

The direction of the field is given by the following *right-hand rule*: If the current were grasped in the right hand with the thumb pointing in the direction of the current, the fingers would indicate the direction of the field.

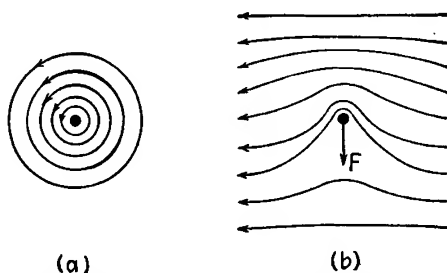


FIG. 11.3.—Magnetic field about a straight current, (a) alone, and (b) in another magnetic field, showing direction of the force upon the current.

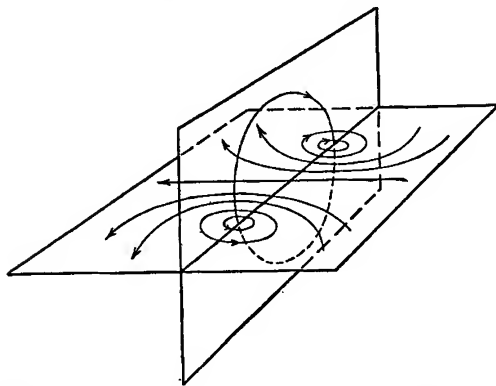


FIG. 11.4.—Magnetic field in a plane containing the axis of a circular current-carrying loop.

The direction of the force acting on a segment of a current-carrying conductor or probe located in a magnetic field, obtained by Fleming's left-hand rule, can also be obtained as follows: Let Fig. 11.3a represent a cross section of the probe carrying a current directed toward the observer. The field of this current is as shown. If the probe is located in a magnetic field perpendicular to the probe, the two fields will combine as in Fig. 11.3b. The force that the

probe experiences can be pictured as a result of the effort of the "crowded" field to "expand."

The magnetic field about a circular current is illustrated in Fig. 11.4, the lines of force being shown in a plane containing the axis of the circular loop. At every point in the plane of the loop the field is perpendicular to this plane. The direction of the field is consistent with the right-hand rule for a straight current element. The magnitude of the field varies at different points along a diameter of the loop. At the axis, the field has the least magnitude within the loop and the magnetizing force is

$$H = \frac{i}{2r} \text{ ampere turns/meter} \quad (11.5)$$

where r is the radius of the loop in meters and i is the current in amperes.

12. Free Electron in Magnetic Field.—The shaded area in Fig. 12.1 represents a cross section of a uniform magnetic field in

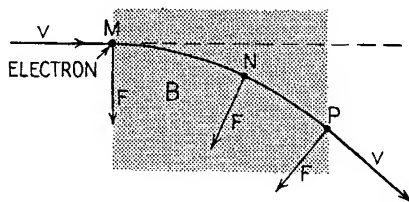


FIG. 12.1.—Deflection of a moving electron in a vacuum in a uniform magnetic field, directed into the paper.

vacuum, with the lines of induction B directed into the paper. When a moving electron enters the field at M , with a velocity v perpendicular to B , it experiences a transverse force

$$F = Bev \text{ newtons} \quad (12.1)$$

where B is the magnetic induction in mks B units (newtons/coulomb/ampere or webers/sq meter), e is the electronic charge in coulombs, and v is the velocity of the electron in meters/sec. The direction of the force is perpendicular to v and is given by Fleming's left-hand rule, Sec. 9, the direction opposite to the velocity v replacing the current i . Under the action of the transverse force F , the direction of motion of the electron changes but always remains perpendicular to the field. The magnitude of the transverse force F and the magnitude of the electron velocity v do not change. In consequence, the electron follows a circular path whose radius is

$$R = \frac{mv}{eB} \text{ meters} \quad (12.2)$$

Numerical values of e and m are given in Sec. 2. The length of time necessary for the electron to complete one excursion around the circumference of the circle is

$$T = \frac{2\pi R}{v} = \frac{2\pi m}{eB} \text{ sec} \quad (12.3)$$

and is independent of the velocity of the electron. However, if the electron leaves the magnetic field as at P , it continues in a straight path in the direction of emergence, the magnitude of its velocity having remained throughout at its initial value.

When an electron moves in vacuum in a direction other than perpendicular to a magnetic field, it experiences a force, perpendicular to both field and motion, that is proportional to the component of the motion perpendicular to the field. No force is exerted on an electron if its motion is collinear with the field.

13. Electromagnetic Induction.—Let Φ be the flux of the magnetic induction B through any electric circuit, for example a loop of conducting wire. A change in the amount of flux Φ gives rise to an emf in the circuit that lasts only as long as the flux is changing. This emf is expressed by

$$\text{Emf} = -N \frac{d\Phi}{dt} \text{ volts} \quad (13.1)$$

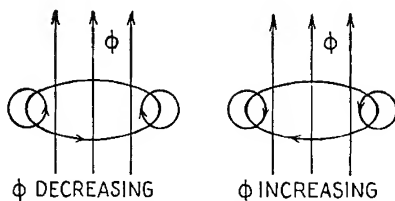


FIG. 13.1.—Direction of induced emf and current in a loop due to change in Φ .

where N is the number of turns linked by the flux Φ .¹ The negative sign indicates that the emf is induced in the circuit in the negative direction when the rate of change of Φ is positive.

The induced emf produces in the closed loop a current having a field as in Fig. 11.4, whose direction is consistent with the right-hand rule. This field is always in such a direction as to aid the original field if Φ is being decreased or to oppose the original field if Φ is being increased (Lenz's law). The sketches in Fig. 13.1 illustrate these principles.

Equation (13.1) has wide application. It does not matter in what way the change in flux linkage is accomplished. The field,

¹ From (13.1) it is seen that the unit of Φ , the weber, equals 1 volt-second.

the circuit, or both may be in motion. If the circuit should already contain a steady emf and current, the induced emf and current are superimposed upon them.

When a current i flows in a closed loop consisting of N turns, there is a magnetic flux Φ through the loop. If the current changes, Φ changes, and there is induced in the circuit an emf expressed by the relation

$$\text{emf} = -N \frac{d\Phi}{dt} = -L \frac{di}{dt} \text{ volts} \quad (13.2)$$

The minus sign indicates that the direction of the emf is such as to oppose the change in the current. The coefficient L is the *self-inductance* of the circuit and may be regarded as a measure of the induced emf at a moment when the current is changing at unit rate. The unit of inductance in the mks system is the *henry*. A 1-volt emf is induced in a circuit of 1-henry inductance when the current in the circuit is changing at the rate of 1 ampere/sec.

When a circuit contains a number of loops in series, so arranged that the flux of any one loop links some or all of the other loops as well, the self-inductance is thereby increased. A coil is thus a circuit element possessing a substantial amount of self-inductance. If the coil is wound on an iron core, its inductance thereby is greatly increased. If a coil is wound on a toroidal core of magnetic material, the magnetic flux Φ within the toroid is μ_r times as great as in the absence of the magnetic material and the inductance L of the coil is μ_r times as great since the external flux usually may be neglected.

When L is constant, the energy stored in a current-carrying circuit is

$$\text{Energy in circuit} = \frac{1}{2} Li^2 \text{ joules} \quad (13.3)$$

where L is in henrys and i in amperes. Thus, half a joule is stored in a 1-henry inductance when the current is one ampere. Compare this equation with (5.5).

All closed electric circuits have a certain amount of self-inductance L , a certain amount of resistance R , and a certain amount of capacitance C . One or two of these quantities may predominate, but all three are always present.

When two circuits are so placed that the flux of induction of one is partly or wholly linked with the other, a change of current in circuit 1 induces an emf in circuit 2, expressed by

$$\text{emf}_2 = -M \frac{di_1}{dt} \text{ volts} \quad (13.4)$$

In like manner, a change of current in circuit 2 induces an emf in circuit 1,

$$\text{emf}_1 = -M \frac{di_2}{dt} \text{ volts} \quad (13.5)$$

The coefficient M is called the *mutual inductance*, "mutual" emphasizing the fact that the coefficients in (13.4) and (13.5) are

TABLE 14.1

Kind	Quantity mks units	Quantity esu	Quantity emu
Length.....	1 meter	100 cm	100 cm
Mass.....	1 kg	1,000 grams	1,000 grams
Time.....	1 sec	1 sec	1 sec
Force.....	1 newton	10^5 dynes	10^5 dynes
Work.....	1 joule	10^7 ergs	10^7 ergs
Power.....	1 watt	10^7 ergs/sec	10^7 ergs/sec
Charge.....	1 coulomb	$3 \cdot 10^9$ statcoulombs	10^{-1} abcoulomb
Current.....	1 ampere	$3 \cdot 10^9$ statamperes	10^{-1} abampere
Electric field.....	1 volt/meter	$\frac{1}{3} \cdot 10^{-4}$ statvolt/cm	10^6 abvolts/cm
Potential difference.....	1 volt	$\frac{1}{300}$ statvolt	10^8 abvolts
Conductivity.....	1 mho/meter	$9 \cdot 10^9$ statmho/cm	10^{-11} abmho/cm
Resistance.....	1 ohm	$\frac{1}{9} \cdot 10^{-11}$ statohm	10^9 abohms
Capacitance.....	1 farad	$9 \cdot 10^{11}$ statfarads or cm	10^{-9} abfarad
Magnetic flux.....	1 weber	$\frac{1}{300}$ esu	10^8 maxwells
Magnetic flux density B	1 weber/sq meter	$\frac{1}{9} \cdot 10^{-6}$ esu/cm ²	10^4 gauss
Magnetic field H	1 ampere turn/meter	$12\pi \cdot 10^7$ esu	$4\pi \cdot 10^{-3}$ oersted
Magnetomotive force.....	1 ampere turn	$12\pi \cdot 10^9$ esu	$4\pi \cdot 10^{-1}$ gilbert
Inductance.....	1 henry	$\frac{1}{9} \cdot 10^{-11}$ stathenry	10^9 abhenry
Permittivity of free space ϵ_0	$[1/(36\pi)] \cdot 10^{-9}$ farad/meter	1	$\frac{1}{9} \cdot 10^{-20}$ (sec/cm) ²
Permeability of free space μ_0	$4\pi \cdot 10^{-7}$ henry/meter	$\frac{1}{9} \cdot 10^{-20}$ (sec/cm) ²	1
Relative permittivity ϵ_r (dielectric-constant).....	Numeric		
Relative permeability μ_r	Numeric		

the same. The unit of mutual inductance in the mks system is the henry, the same unit as for self-inductance.

Positive direction of current having been arbitrarily designated in each circuit, if the flux linked with circuit 1 due to positive current in circuit 2 aids the flux due to positive current in circuit 1, M is numerically positive. If these fluxes oppose, M is numerically negative (see Chap. I, Sec. 11).

14. Comparison of Units.—Table 14.1 gives a comparison of the mks system of units, the electrostatic units (esu) and the electromagnetic units (emu). The comparison is numerical where the units are dimensionally different, *i.e.*, involve mass, length, time in different degrees. One mks unit of the quantity in the first column is expressed in the appropriate number of electrostatic units and electromagnetic units. For example, 1 mks unit of current (1 ampere) equals $3 \cdot 10^9$ esu of current ($3 \cdot 10^9$ statamperes), or 10^{-1} emu of current (10^{-1} abamperes). The relative permittivity or dielectric constant is a numeric of the same magnitude in all systems. In the table, the characteristic velocity of electromagnetic waves in free space is taken as $3 \cdot 10^8$ meters/sec or 300 meters/microsec.

APPENDIX C

DESCRIPTION OF A DIRECT-CURRENT SOURCE IN TERMS OF ELECTRICAL CONSTANTS

1. Constant-voltage and Constant-current Representation.—

Figure 1.1a indicates a source of electrical energy connected to a dissipative load R_L . The action of the source upon the load may be described in terms of electrical constants if its terminal potential varies linearly with the load current I_L . In Fig. 1.2b the solid diagonal line represents this characteristic plotted for a particular d-c generator. At zero current the terminal potential is 40 volts. This voltage at zero current is a significant constant of the generator

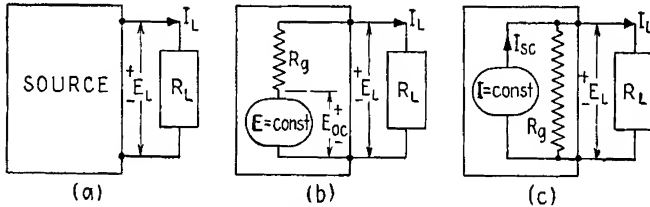


FIG. 1.1.—Alternative descriptions (b) and (c) of a d-c generator (a).

and may be described as its *open-circuit voltage* E_{oc} or its emf. For the generator of this particular example, the terminal voltage decreases linearly as the load current increases, Fig. 1.2b, becoming zero when the load current is 4 ma. This value of current at zero voltage is likewise a significant constant of the generator and occurs when the load resistance is zero, or when the generator terminals are short-circuited; it is called the *short-circuit current* I_{sc} . The equation for the generator characteristic may be written in terms of these two intercepts on the axes:

$$E_L = E_{oc} - \frac{E_{oc}}{I_{sc}} I_L \quad (1.1)$$

The ratio E_{oc}/I_{sc} has the dimension of ohms and is defined as the generator resistance $R_g = E_{oc}/I_{sc}$. It is equal to the decrease in

terminal voltage per ampere of current through the load and is determined by the slope of the line of Fig. 1.2b. For the generator shown, $R_g = 10,000$ ohms.

The load characteristic of the generator also can be described in terms of R_g in two other forms,

$$E_L = E_{oc} - R_g I_L \quad \text{or} \quad E_L = (I_{sc} - I_L) R_g \quad (1.2)$$

The equation at the left indicates that the d-c source may be represented by a zero-resistance generator of constant voltage E_{oc} in series with an internal resistance R_g , Fig. 1.1b. The decrease of E_L with increasing I_L thus is accounted for by the increasing IR drop in R_g . The right-hand equation (1.2) suggests the circuit of Fig.

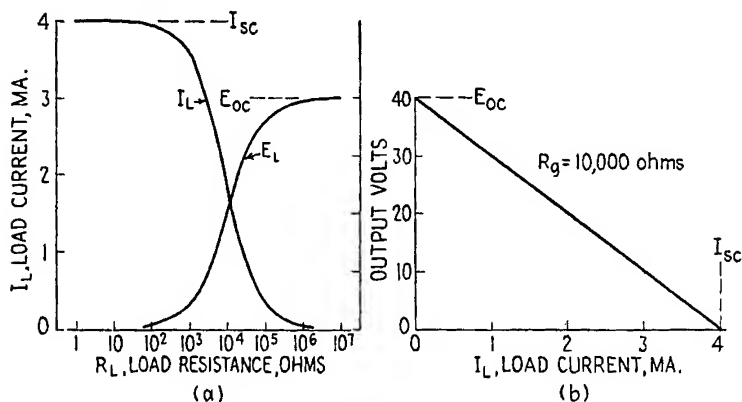


FIG. 1.2.—Load characteristics of a particular d-c generator.

1.1c, in which an infinite-resistance generator, delivering a constant current I_{sc} , is in parallel with the same internal resistance R_g . Here the increase in load current causes a decrease in the current through R_g and therefore a reduced terminal potential. The two descriptions differ in their predictions as to the energy dissipated in the generator, but they are equivalent insofar as load currents and voltages are concerned. Which equivalent circuit is more convenient depends upon the conditions of the problem.

If R_g , the resistance of the generator studied, had been 100 ohms instead of 10,000 (E_{oc} remaining the same), the E_L - I_L characteristic would be given in part by the almost horizontal dashed line of Fig. 1.3. Over the range of currents shown (which correspond to load resistances much larger than 100 ohms) the percentage change from the open-circuit voltage is negligible, and within this range

the generator approximates a constant-voltage device. Under these circumstances the generator is conveniently described by the single constant E_{oc} , and the series resistor R_g may be omitted from the diagram of Fig. 1.1b. If the generator resistance R_g had been 1 megohm (I_{sc} remaining the same), the almost vertical dashed line at the right in Fig. 1.3 would describe its behavior. In this range of voltage variation, corresponding to values of load resistance that are very small in comparison with 1 megohm, the source approximates a constant-current generator. For this condition, resistor R_g in Fig. 1.1c may be neglected and the generator described by the single constant I_{sc} .

The same source may approximate either a constant-current or a constant-voltage generator depending upon the ratio of the load

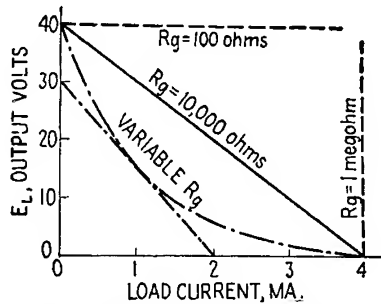


FIG. 1.3.—Load characteristics of generators having different internal resistances R_g .

resistance to R_g , the internal resistance of the generator. In Fig. 1.2a the variation of E_L and I_L is shown for the generator previously described whose R_g was 10,000 ohms. As the load resistance is increased from a value of 1 ohm, the current I_L remains almost constant at its short-circuit value of 4 ma. At a load resistance of 100 ohms (1 per cent of R_g) the load current drops about 1 per cent, and at 1,000 ohms (10 per cent of R_g) the decrease is less than 10 per cent. For many purposes the generator may be represented as a constant-current generator in this range of loads. As the load resistance increases to a value equal to R_g , the load current drops to one-half its short-circuit value, and the terminal voltage E_L rises to one-half its open-circuit value. Further increase of R_L to $10R_g$ and $100R_g$ raises the generator voltage to within 10 per cent and 1 per cent, respectively, of the open-circuit voltage. Above these values of load resistance, constant-voltage operation is a good approximation.

In the range of resistance loads between constant-voltage and

constant-current operation, the representation used may be determined by the nature of the load. If it consists only of series elements, the constant-voltage representation is more convenient, as the generator resistance adds only one more series element. If the load consists of parallel branches, the constant-current description adds R_g as an additional parallel element.

2. Nonlinear Characteristics.—Many actual generators do not yield a straight E_L - I_L characteristic, but instead one similar to the dot-dash curve in Fig. 1.3. The action of this generator upon a load cannot be described in terms of a fixed emf and fixed R_g for all values of load current. If the load current does not change greatly and the curvature of the characteristic is small, approximate values of emf and R_g may be established for a limited range of currents. For example, the portion of the dot-dash curve of Fig. 1.3 in the neighborhood of 1 ma may be simulated by the tangent to the curve at that point, so that for currents in the neighborhood of 1 ma the generator may be simulated by a fictitious generator having values of $E_{oc} = 30$ volts, $I_{sc} = 2$ ma, and $R_g = 15,000$ ohms.

APPENDIX D

SYMBOLS AND ABBREVIATIONS

The purpose of this appendix is to explain the general plan underlying the symbols used in this text and to list certain symbols and abbreviations. Symbols for impedance, resistance, inductance, capacitance, etc., are conventional; these and the subscripts used with them are explained in the text of each chapter. The explanations given below are concerned mainly with currents and voltages.

1. Capital Letters.—Capital, or upper-case, letters (when used for currents and voltages) are used for steady, or d-c, components, for repetitive a-c and variational components, and in general for currents and voltages that continually recur cyclically or systematically.

Capital letters are also given special meanings as indicated by the text of the chapter concerned.

2. Lower-case Letters.—Lower-case, or small, letters are used to indicate instantaneous values of currents and voltages. Both capital and lower-case letters used as symbols are printed in italics to distinguish them from the accompanying text.

3. Use and Omission of Diacritical Marks.—Two diacritical marks are used extensively in this text—a short bar and a short convex arc, or curved line. These are placed over capital italic letters.

The bar denotes steady, or average, components, or d-c values. Thus, $\bar{E} = 10$ volts means that the voltage in question is a d-c, or steady, voltage, existing alone or as a component of a more complex waveform, and has a value of +10 volts in accordance with the polarity assumed in the text.

The arc is used over a capital letter to simulate one half cycle of a sinusoid and denotes the amplitude of a sinusoidal component. Thus, $\hat{E} = 4 + j3 = 5/36.9^\circ$ volts indicates a voltage of sinusoidal waveform whose amplitude is 5 volts, having rectangular components also of sinusoidal waveform whose amplitudes are 4 and 3 volts.

Where no diacritical mark is used, the capital italic letter indicates rms value. Thus, $E = 2.83 + j2.12 = 3.54/36.9^\circ$ volts expresses in terms of rms values the same voltage as in the previous paragraph.

As another example, suppose the instantaneous value of a certain current in milliamperes (abbreviated ma) is given by

$$i = 10 + 5 \sin (\omega_1 t + \theta_1) + 3 \sin (\omega_2 t + \theta_2)$$

This is denoted in symbols by

$$i = \bar{I} + |\hat{I}_1| \sin (\omega_1 t + \theta_1) + |\hat{I}_2| \sin (\omega_2 t + \theta_2)$$

where \bar{I} denotes the d-c component and $|\hat{I}_1|$ and $|\hat{I}_2|$ the magnitude values of the amplitudes of the a-c components. In accordance with the notation, $\bar{I} = 10$ ma; $|\hat{I}_1| = 5$ ma; $|I_1| = 3.54$ ma; $|\hat{I}_2| = 3$ ma; $|I_2| = 2.12$ ma. The rms value of i is denoted by I ; for this example, $I = 11.58$ ma.

A diacritical mark used less extensively in this text is the tilde (\sim). This mark approximates in form one cycle of a sinusoid and is used here to denote quantities involving alternating current. Suppose that a circuit has a certain value of resistance to direct current and another value to alternating current, the frequency being specified where necessary. This condition exists where the resistance of an iron-core coil is involved. The resistance to direct current would be denoted by \bar{R} and the resistance to the alternating current by \tilde{R} . Also, the power dissipation in a circuit resulting from the direct current would be designated by \bar{P} and that resulting from the alternating current by \tilde{P} . The tilde is used only where it is necessary to distinguish between d-c and a-c values.

The diacritical mark \wedge placed over a letter signifies peak, or maximum, value. For a sinusoid, peak and amplitude values are equal.

4. Vacuum-tube Voltages and Currents.—This section catalogues most of the symbols used for vacuum-tube voltages and currents.

The subscripts b, c, d, z refer, respectively, to the plate, grid, screen grid, plate-load impedance. The subscripts p, g, L refer respectively to variational components concerned with the plate, grid, plate-load impedance. The subscript 0 (zero) when used with a symbol denoting average value means the quiescent value; when used with a lower-case letter the subscript 0 means variational value measured *from* the quiescent value.

Symbol	Meaning
e_b	Total instantaneous voltage of plate with respect to cathode

Symbol	Meaning
e_{bb}	Instantaneous value of plate-supply voltage
e_c	Total instantaneous voltage of grid with respect to cathode
e_{cc}	Instantaneous value of grid-supply voltage
e_d	Total instantaneous voltage of screen grid with respect to cathode
e_g	Instantaneous value of variational component of grid voltage with respect to cathode, measured from the average value (is a-c part of e_c , has no d-c component)
e_{g0}	Instantaneous value of variational component of grid voltage with respect to cathode, measured from the quiescent value (equals e_c minus \bar{E}_{c0} , may have a d-c component)
e_L	Instantaneous value of variational component of voltage drop across the load in the direction of positive plate current, measured from the average value (is a-c part of e_z , has no d-c component)
e_{L0}	Instantaneous value of variational component of voltage drop across the load in the direction of positive plate current, measured from the quiescent value (equals e_z minus \bar{E}_{z0} , may have a d-c component)
e_p	Instantaneous value of variational component of plate voltage with respect to cathode, measured from the average value (is a-c part of e_b , has no d-c component)
e_{p0}	Instantaneous value of variational component of plate voltage with respect to cathode, measured from the quiescent value (equals e_b minus \bar{E}_{b0} , may have a d-c component)
e_z	Total instantaneous voltage drop across the plate load in the direction of positive plate current
\bar{E}_b	Average voltage of plate with respect to cathode
\bar{E}_{bb}	Average value of plate-supply voltage
\bar{E}_{b0}	Quiescent voltage of plate with respect to cathode (is quiescent value of E_b or of e_b)
\bar{E}_c	Average voltage of grid with respect to cathode
\bar{E}_{cc}	Average value of grid-supply voltage
\bar{E}_{c0}	Quiescent grid voltage with respect to cathode (is quiescent value of \bar{E}_c or of e_c)
\bar{E}_d	Average voltage of screen grid with respect to cathode
\bar{E}_{dd}	Average value of screen-grid supply voltage
E_g	Rms value of a-c components of grid voltage with respect to cathode (is rms value of e_g)
E_L	Rms value of a-c components of voltage drop across the plate load in the direction of positive plate current (is rms value of e_L)
E_p	Rms value of a-c components of plate voltage with respect to cathode (is rms value of e_p)
\bar{E}_z	Average voltage drop across the plate load in the direction of positive plate current (is average value of e_z)
\bar{E}_{z0}	Quiescent voltage drop across the plate load in the direction of positive plate current (is quiescent value of \bar{E}_z or of e_z)
i_b	Total instantaneous plate current to the plate through the plate terminal
i_{bb}	Total instantaneous plate-supply current
i_c	Total instantaneous grid current to the grid through the grid terminal

Symbol	Meaning
i_d	Total instantaneous screen-grid current to the screen grid through the screen-grid terminal
i_g	Instantaneous value of variational component of grid current, measured from average value (is a-c part of i_c , has no d-c component)
i_{g0}	Instantaneous value of variational component of grid current, measured from the quiescent value (equals i_c minus \bar{I}_{c0} , may have a d-c component)
i_p	Instantaneous value of the variational component of current, measured from the average value (is a-c part of \bar{I}_b , has no d-c component)
i_{p0}	Instantaneous value of variational component of plate current, measured from the quiescent value (equals i_b minus \bar{I}_{b0} , may have a d-c component)
\bar{I}_b	Average plate current (is average value of i_b)
\bar{I}_{bb}	Average value of plate-supply current
\bar{I}_{b0}	Quiescent plate current (is quiescent value of \bar{I}_b or of i_b)
\bar{I}_c	Average grid current (is average value of i_c)
\bar{I}_{c0}	Quiescent grid current (is quiescent value of \bar{I}_c or of i_c)
\bar{I}_d	Average screen-grid current (is average value of i_d)
I_g	Rms value of a-c components of grid current (is rms value of i_g)
I_p	Rms value of a-c components of plate current (is rms value of i_p)
\bar{I}_z	Average value of current through the plate load

5. Abbreviations.—The following is a list of abbreviations commonly used in this text:

Abbreviation	Meaning
a-c	Alternating-current, alternating (<i>e.g.</i> , a-c component)
a-f	Audio-frequency
AVC	Automatic volume control
cps	Cycles per second
CRO	Cathode-ray oscillograph
d-c	Direct-current, steady or constant (<i>e.g.</i> , d-c component)
emf	Electromotive force
i-f	Intermediate-frequency
kcps	Kilocycles per second (thousands of cycles per second)
ma	Milliampere
max	Maximum
mcps	Megacycles per second (millions of cycles per second)
mh	Millihenry
r-f	Radio-frequency
RFC	Radio-frequency choke coil
rps	Revolutions per second
sec	Second
VTVM	Vacuum-tube voltmeter
μ a	Microampere
μ f	Microfarad
μ h	Microhenry

Abbreviations		Meaning
μs	Microsecond	
$\mu\mu f$	Micromicrofarad (10^{-12} farad)	

6. Mathematical Symbols.—The mathematical symbols of this text are the usual ones. A few are given here.

Symbol		Meaning
		Absolute magnitude of the quantity between the bars
\doteq		Is approximately equal to, equals approximately
$>$		Is greater than
$>>$		Is very much greater than (approximately ten times or more)
$<$		Is less than
$<<$		Is very much less than (approximately one-tenth or less)
∞		Infinity
\propto		Is proportional to
\sphericalangle		At the angle

PROBLEMS

Chapter I

1. The function shown in Fig. 1 is periodic with a period of 5. Over any whole number of periods, find: (a) the average value of the function; (b) the rms value. *Ans.:* (a) 3.00; (b) 4.47.

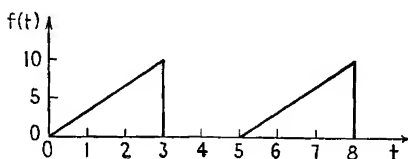


FIG. 1.

2. A sinusoidal emf is given by the equation $e = 196 \sin (10,000t - \frac{1}{3}\pi)$, where e is in volts and t is in seconds. (a) Express the rms value of this voltage in polar form. (b) Express this voltage in rectangular complex form. (c) What is the frequency of this emf? (d) What is the first time after $t = 0$ when $e = 98.0$ volts? *Ans.:* (a) $139/\underline{-60^\circ}$ volts; (b) $69.3 - j120$ volts; (c) 1,590 cps; (d) $1.57 \cdot 10^{-4}$ sec.

3. A current is given in milliamperes by the formula $i = 5.00 + 10.00 \sin 1,000t$. (a) What is the average value of this current over any whole number of cycles? (b) What is the rms value of the current? (c) If the current is "rectified" so that all negative parts of the current curve are inverted to become positive, determine the average value over any whole number of periods. *Ans.:* (a) 5.00 ma; (b) 8.66 ma; (c) 7.18 ma.

4. (a) A voltage $e = 12.5 \sin (20,000t - 20.0^\circ)$ is applied to a pure inductance of 1.30 mh. Write the sine wave equation of the current through the inductance. (b) If the same voltage is applied to a pure capacitance of $800 \mu\text{f}$, write the sine wave equation of the current through the capacitance. *Ans.:* (a) $i = 0.481 \sin (20,000t + 70.0^\circ)$ ampere; (b) $i = 2.00 \cdot 10^{-4} \sin (20,000t - 110^\circ)$ ampere.

5. (a) Figure 5 represents the voltage in volts applied to a pure capacitance of $5 \mu\text{f}$. Sketch the waveform of the current through the capacitance for the interval shown giving numerical values. (b) If Fig. 5 represents the current in milliamperes through a pure inductance of 2 henrys, sketch the waveform of the voltage across the inductance giving numerical values.

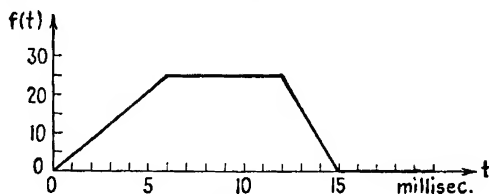


FIG. 5.

6. A voltage $e = 50.0 \sin 800t$ volts is applied across a pure resistance of 20.0 ohms. (a) What is the instantaneous power at $t = 0.001$ sec? (b) What is the average power dissipated in the resistor? *Ans.*: (a) 64.3 watts; (b) 62.5 watts.

7. A series circuit contains a pure resistance of 7.50 ohms, a pure inductance of $175 \mu\text{h}$, and a pure capacitance of $325 \mu\text{f}$. A voltage $e = 7.50 \sin 5.00 \cdot 10^4 t$ volts is applied across the series circuit. Find: (a) the impedance in rectangular complex form; (b) the impedance in polar form; (c) the power factor; (d) the current in polar form; (e) the power. *Ans.*: (a) $7.50 + j260$ ohms; (b) $260/88.35^\circ$ ohms; (c) 0.288; (d) $0.0204/-88.35^\circ$ ampere; (e) 3.13 mw.

8. In the circuit of Fig. 8, $R_L = 15.0$ ohms, $L = 75.0 \mu\text{h}$, $R_C = 50.0$ ohms, $C = 40.0 \mu\text{f}$, $R = 1,000$ ohms. At a frequency of 2,750 kcps, find: (a) the absolute value of the total conductance; (b) the absolute value of the total susceptance; (c) the total admittance in rectangular complex form; (d) the total admittance in polar form; (e) the power factor of the circuit. *Ans.*: (a) $1,033 \mu\text{mhos}$; (b) $81.2 \mu\text{mhos}$; (c) $1,033 - j81.2 \mu\text{mhos}$; (d) $1,036/-4.50^\circ \mu\text{mhos}$; (e) 0.997.

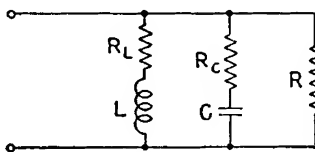


FIG. 8.

9. Figure 9 represents a voltage divider. The input voltage E_1 has an rms value of 10.0 volts at a frequency of 1,000 cps. C has a capacitance of $0.05 \mu\text{f}$. The output voltage E_2 is to have an rms value of 7.50 volts. Determine the necessary value of R . *Ans.*: 1,062 ohms.

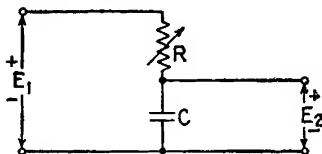


FIG. 9.

10. Figure 10 represents a phase-shifting circuit. R and C are always adjusted so that the magnitude of their total series impedance is 5,000 ohms. The input voltage has a frequency of 800 cps. (a) What must be the value of R and C to produce a phase shift of 30° between the input voltage E_1 and the output voltage E_2 ? (b) Does the voltage E_2 lag or lead E_1 by 30° ? *Ans.*: (a) $R = 2,500$ ohms, $C = 0.0459 \mu\text{f}$; (b) lags.

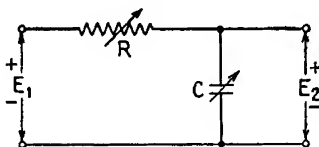


FIG. 10.

Chapter II

1. A series circuit consists of 12.0 ohms of resistance, 180 μh of inductance, and 250 μf of capacitance. A source of constant voltage of 5.00 volts rms at variable frequency and zero phase angle is connected across the circuit. Determine: (a) the resonant frequency of the circuit; (b) the rms value of the current at the resonant frequency; (c) the average value of the total energy stored in the inductance and capacitance at resonance; (d) Q_r ; (e) the rms voltage across the capacitance at resonance; (f) the rms voltage across the inductance at resonance. *Ans.:* (a) 750 keps; (b) 0.417 ampere; (c) $3.125 \cdot 10^{-6}$ joule; (d) 70.7; (e) 354 volts; (f) 354 volts.

2. A series circuit consisting of 4.10 ohms of resistance, 81.0 μh of inductance, and 225 μf of capacitance is excited by a constant-voltage generator of variable frequency. (a) What is the resonant frequency of the circuit? (b) At what frequency is maximum power delivered to the circuit? (c) What is Q_r for the circuit? (d) At what frequencies will one-half the maximum power be delivered? *Ans.:* (a) 1,179 keps; (b) 1,179 keps; (c) 146; (d) 1,175 and 1,183 keps.

3. At a frequency of 1,400 keps, coil *A* has a resistance of 9.04 ohms and an inductance of 185 μh , while coil *B* has a resistance of 9.49 ohms and an inductance of 205 μh . If each of these coils were to be tuned to series resonance at 1,400 keps by a capacitor of negligible losses, which coil would form the most selective circuit? *Ans.:* Coil *B*.

4. Table 4 shows the variation of current through a series circuit of R , L , and C when excited by a constant-voltage generator at a variable frequency. The generator delivers 5.00 volts rms. The maximum current is 344 ma. Find: (a) Q_r ; (b) R ; (c) L ; (d) C . *Ans.:* (a) 55.2; (b) 14.5 ohms; (c) 145 μh ; (d) 224 μf .

TABLE 4

f keps	I ma
870	172
876	243
879	292
884	344
889	292
892	243
898	172

5. In the circuit of Fig. 5, R is 200 ohms, L is 10.0 mh. The circuit is excited by a constant-current generator which supplies 1.00 ma at the resonant frequency of the circuit. If the resonant frequency is 159 keps, determine: (a) Q_r ; (b) the value of C ; (c) the impedance of the circuit at resonance; (d) the voltage across C at resonance; (e) the current through L at resonance. *Ans.:* (a) 50; (b) 100 μf ; (c) 500,000 ohms; (d) 500 volts; (e) 50 ma.

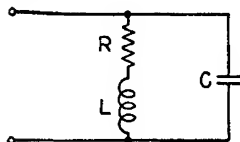


FIG. 5.

6. In the circuit of Fig. 5, R is 6.24 ohms, L is 210 μh , and C is 156 μf . Find: (a) the impedance of the circuit at parallel resonance; (b) the impedance

5.24 kcps below the resonant frequency; (c) the phase angle of the total current with respect to the applied voltage under the conditions of (b). *Ans.*: (a) 216,000 ohms; (b) 88,700 ohms; (c) 65.7° lagging.

7. In the circuit of Fig. 5, R is 15.0 ohms, L is $125\ \mu\text{h}$, C is $45.0\ \mu\text{f}$. (a) What is Q_r for the circuit? (b) At what frequency is the circuit impedance greatest? (c) At what frequencies is the circuit impedance 0.707 times the maximum impedance? *Ans.*: (a) 111; (b) 2.12 meps; (c) 2.111 and 2.129 meps.

8. In the circuit of Fig. 5, R is 4.80 ohms, L is $88.0\ \mu\text{h}$, C is $375\ \mu\text{f}$. Determine: (a) Q_r ; (b) the resonant frequency of the circuit. (c) If the circuit were loaded by connecting a pure resistance of 16,000 ohms across it, what would be the new value of Q_r ? *Ans.*: (a) 101; (b) 876 kpcs; (c) 25.

9. In the circuit of Fig. 9, R is 15.0 ohms, L is $95.0\ \mu\text{h}$, the generator resistance R_g is 8,000 ohms. The circuit of R , L , C_1 , C_2 is resonant at a frequency of 6.50 meps. What should be the values of C_1 and C_2 to make the input impedance of the tuned circuit equal to the generator impedance? *Ans.*: $C_1 = 6.93\ \mu\text{f}$; $C_2 = 70.7\ \mu\text{f}$.

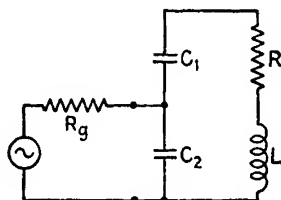


FIG. 9.

10. The circuit of Fig. 10 is to be adjusted to have the total current a maximum for a frequency of 1.0 meps and a minimum for a frequency of 1.2 meps. Find the values of C_1 and C_2 which will produce these conditions. *Ans.*: $C_2 = 85.4\ \mu\text{f}$; $C_1 = 38.5\ \mu\text{f}$.

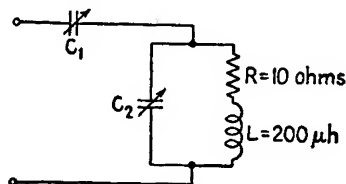


FIG. 10.

Chapter IV

1. Show that the equations for balance in the Campbell-Heaviside bridge, Fig. 1, are

$$R_1 R_4 = R_2 R_3 \quad (1)$$

$$L_3 = \frac{R_1}{R_2} (L_4 \pm M) \pm M \quad (2)$$

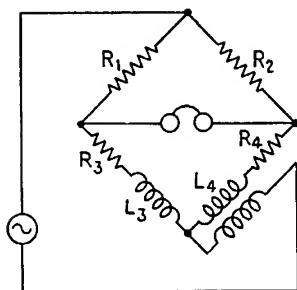


FIG. 1.

2. It is desired to obtain the effective resistance R , inductance L , and Q for a coil at a frequency of 6.50 mcps. The coil is connected in series with a calibrated standard capacitor of negligible losses, and the series circuit is connected across a constant-voltage generator operating at 6.50 mcps. As the capacitor is varied, the current through the circuit is observed. Maximum current occurs when the capacitance is $12.56 \mu\text{mf}$, and 0.707 times the maximum current is reached when the capacitance is either 12.47 or $12.64 \mu\text{mf}$. For the coil, find (a) Q ; (b) R ; (c) L . *Ans.:* (a) 148; (b) 13.2 ohms; (c) $47.8 \mu\text{h}$.

Chapter V

1. Figure 1 represents a "lattice" network, $abcd$, connected to a 3-ohm load. The values given are resistances in ohms. (a) Find the input resistance R_{ab} with the 3-ohm load connected as shown. (Hint: unfold the triangle bcd about the line cd to form a bridge network, and then calculate the equivalent T of one of the triangular meshes.) (b) Calculate R_{O1} , R_{S1} , R_{O2} , R_{S2} for the network $abcd$, the 3-ohm load being disconnected. (c) Calculate the arms of the T network equivalent to the network $abcd$. *Ans.*: (a) 8.77 ohms; (b) 14.64, 7.50, 13.80, 7.06 ohms; (c) $R_1 = 4.70$, $R_2 = 3.87$, $R_3 = 9.93$ ohms.

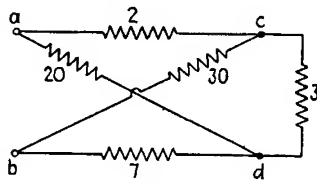


FIG. 1.

2. The open-circuit impedances Z_{O1} , Z_{O2} , and the short-circuit impedance Z_{S1} of a certain network are $Z_{O1} = 500(1 - j10)$; $Z_{O2} = 400(1 - j10)$; $Z_{S1} = 250(1 - j10)$. Find: (a) the short-circuit impedance Z_{S2} ; (b) the impedances of the equivalent T network; (c) the impedances of the equivalent Π network. *Ans.*: (a) $Z_{S2} = 200(1 - j10)$; (b) $Z_1 = 184(1 - j10)$, $Z_2 = 84(1 - j10)$; $Z_3 = 316(1 - j10)$; (c) $Z_A = 1,192(1 - j10)$, $Z_B = 317(1 - j10)$, $Z_C = 544(1 - j10)$.

3. At the output terminals of a certain network, the open-circuit voltage is 100 volts and the internal impedance is $4 - j6$ ohms. (a) What is the maximum power the network could deliver to a load? What current would flow? (b) What power would be delivered to the load if the load impedance were the complex conjugate of that used in (a)? What current would flow? *Ans.*: (a) 625 watts, $12.5/0^\circ$ amperes; (b) 192.1 watts, $6.93/56.3^\circ$ amperes.

4. What series load circuit would take the maximum power from the active network of Fig. 4 if the load were connected between a and b ? $E = 4.00/0^\circ$ volts at 1,600 cps, $R = 900$ ohms, $L = 0.0750$ henry, $C = 0.0500 \mu\text{f}$. *Ans.*: series circuit of 748 ohms resistance, $0.239 \mu\text{f}$ capacitance.

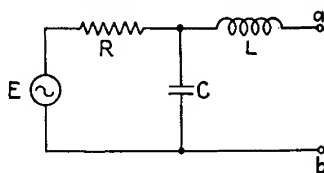


FIG. 4.

5. Figure 5 represents a network designed to match a load R_L of 15,000 ohms to a generator whose internal resistance is 10,000 ohms, at a frequency for which $\omega = 10^6$. What are the values of L and C which permit the maximum transfer of power from the generator to the load? *Ans.*: $L = 7.07$ mh, $C = 47.1$ μ f.

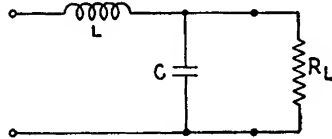


FIG. 5.

6. The plate circuit of a triode consists of two coils and a capacitor connected as in Fig. 6. The mutual inductance between coils is negligible, and the taps shown in the figure are moved in opposite directions so that the sum of the two inductances is kept constant at $L = 50$ μ h, and the sum of the two resistances of the coils is kept constant at $R = 10$ ohms. (a) Determine C and the position of the taps in order to make the a-c impedance of the load equal to a resistance of 20,000 ohms at $\omega = 10^7$. (b) What is the highest resistance this circuit can have at $\omega = 10^7$? *Ans.*: (a) $C = 200$ μ f, $L' = 5.3$ μ h, $L'' = 44.7$ μ h; (b) 25,000 ohms.

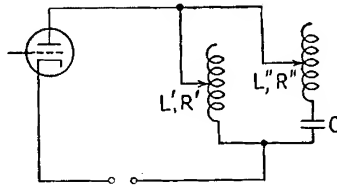


FIG. 6.

7. In Fig. 7, $E = 10/\sqrt{2}$ volts rms, $f = 1$ mcps, $R_1 = 50$ ohms, $C = 1,000$ μ f, $R_2 = 100$ ohms. Find the current in R_2 by using Thévenin's theorem. *Ans.*: $65.3/-11.82^\circ$ ma rms.

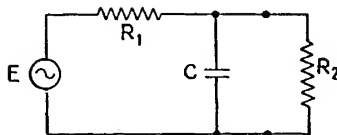


FIG. 7.

8. Figure 8 represents an audio transformer whose secondary winding is connected to a pure resistance load of 8 ohms. For the primary, L_1 equals 50 henrys, R_1 equals 120 ohms; and for the secondary, L_2 equals 0.1 henry and R_2 equals 0.1 ohm. Assume that M^2 equals $L_1 L_2$. (a) What is the input imped-

ance of the transformer at 400 cps? (b) What is the turns ratio of the transformer? *Ans.*: (a) $4,166 + j125.6$ ohms; (b) 22.4.

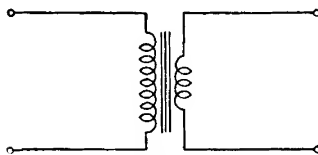


FIG. 8.

9. The primary and secondary resistances and inductances of a transformer are measured and found to be: $R_1 = 580$ ohms; $L_1 = 68$ henrys; $R_2 = 1$ ohm; $L_2 = 50$ mh. Assume $k = 1$ and that shunt capacitances are negligible. (a) What is the impedance at 400 cps, looking into the primary, with an 8-ohm speaker (assume pure resistance) connected across the secondary? (b) At what frequency is this transformer no longer satisfactory for connecting the speaker to a tube whose load resistance should be 11,000 ohms, plus or minus 10 per cent? *Ans.*: (a) $12,760 + j872$ ohms; (b) 51.2 cps.

10. For measurement of the high-frequency current in an r-f coil, an r-f ammeter of the thermocouple type is used with a loop or pickup coil connected across its terminals. For full deflection, a current of 120 ma in the pickup coil is needed. The total resistance in the pickup coil circuit is 1 ohm; the inductance of the pickup coil is $2 \mu\text{h}$, and the inductance of the r-f coil is $200 \mu\text{h}$. What coefficient of coupling is required if a current of 3 amperes in the r-f coil is to give full-scale deflection of the r-f ammeter? Assume the frequency to be greater than 1 mcps. *Ans.*: $k = 0.4$ per cent.

11. In the matching network of Fig. 11, $L = 20 \mu\text{h}$, $C = 0.005 \mu\text{f}$, $R_L = 50$ ohms. (a) At what frequency is the input impedance purely resistive? (b) What is the input resistance at that frequency? *Ans.*: (a) 822 keps; (b) 80 ohms.

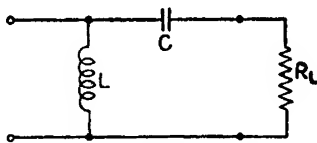


FIG. 11.

12. In the circuit of Fig. 12, $1/(\omega C_2)$ is made equal to ωL_2 . The capacitor C_1 is then adjusted so that the input impedance is a pure resistance. What is the magnitude of this impedance if $R_1 = R_2 = 10$ ohms, $M = 10 \mu\text{h}$, and $f = 10^6$ cps? *Ans.*: 405 ohms.

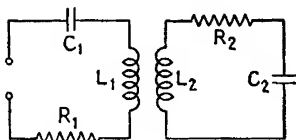


FIG. 12.

Chapter VI

1. A capacitor of $10,000\ \mu\text{f}$ capacitance in series with a resistance of 8 ohms is suddenly connected to a source of 2,000 volts. (a) How long will it take for the charge on the capacitor to be within 0.1 per cent of its final value? (b) What is the final value of the charge? (c) What is the initial value of the current? *Ans.:* (a) $0.553\ \mu\text{sec}$; (b) 20 microcoulombs; (c) 250 amperes.

2. A 1,000-cycle square-wave voltage, varying from -1 volt to $+1$ volt, is impressed on a circuit containing R and C in series. Plot on one graph the applied voltage, the voltage across C ; and the voltage across R for the following cases: (a) $R = 5,000$ ohms, $C = 0.1\ \mu\text{f}$; (b) $R = 100,000$ ohms, $C = 0.1\ \mu\text{f}$; (c) $R = 1,000$ ohms, $C = 0.01\ \mu\text{f}$. It is sufficient to calculate only enough data to show the form of the curves. Assume that the square wave has been applied long enough for the transients to repeat every cycle.

3. A square-wave voltage of 50 cps, varying from -1 volt to $+1$ volt, is applied to a circuit consisting of $R = 500$ ohms and $L = 1$ henry connected in series. (a) Find the time constant. (b) Determine the peak values of the voltages across the resistance and the inductance. (c) Plot the waveform of the voltage across the inductance L , assuming that the square-wave voltage has been applied for many cycles. *Ans.:* (a) 0.002 sec; (b) $E_{R\ \text{peak}} = 0.987$, $E_{L\ \text{peak}} = 1.987$.

4. The circuit shown in Fig. 4 has the following voltage and component values: $\bar{E} = 100$ volts, $R_1 = 20$ ohms, $R_2 = 1,000$ ohms, $L = 10$ henrys, $R_L = 100$ ohms. Find: (a) the time constant for the transients that occur during the time the switch S is closed; (b) the time constant for the transients that occur when the switch S is opened; (c) the maximum voltage that occurs across R_2 when the switch S is opened after being closed for a long time. *Ans.:* (a) 0.0836 sec; (b) 0.00909 sec; (c) 820 volts.

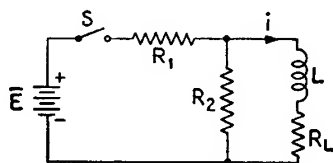


FIG. 4.

5. A capacitor of $C = 0.25\ \mu\text{f}$ is first charged to 2,000 volts. It is then discharged through a coil of $L = 0.1$ henry and $R = 5$ ohms. (a) What is the frequency of the oscillatory discharge? (b) What is the logarithmic decrement? (c) Find the time after the initiation of the discharge at which the charge on the capacitor first reaches zero. Compare with a quarter period of the oscillation. (d) Draw a graph of the charge q on the capacitor plates as a function of time. *Ans.:* (a) 1,007 cps; (b) 0.0248 neper; (c) $247.7\ \mu\text{sec}$.

6. A square-wave 60-cps voltage is applied as indicated in Fig. 6. (a)

Plot the current as a function of time for a complete cycle of the square wave if $(R^2/4L^2)$ is less than $(1/LC)$ and if the frequency of oscillation is very much greater than 60 cps. Assume that the current reaches its final value before the occurrence of a voltage reversal of the square wave. (b) Plot the corresponding curve of charge on the capacitor, indicating coincidence in time for the curve of charge and the curve of current.

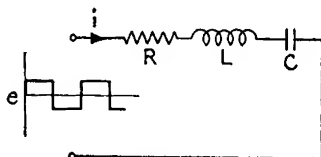


FIG. 6.

7. The switch S in Fig. 7 is closed for $1/10,000$ of a second and opened again. (a) Plot the voltage across the capacitor as a function of time during the interval the switch is closed, indicating the time scale. (b) Plot the voltage across the capacitor from the opening of the switch, using a different time scale.

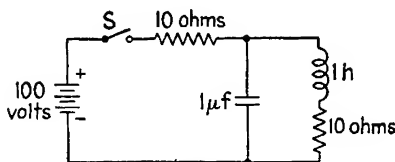


FIG. 7.

8. A surge generator has 20 capacitors of $0.25 \mu\text{f}$ each. The circuit is arranged so that they are all charged in parallel to 100,000 volts and then are connected in series and discharged through a circuit composed of a resistance of 260 ohms and an inductance of $200 \mu\text{h}$. (a) What type of discharge takes place and why? (b) If the resistance of the circuit is adjusted to the critical value, calculate the time required for the discharge current to reach the maximum value. What is the maximum current? *Ans.:* (b) $1.581 \mu\text{sec}$, 5,820 amperes.

9. For the circuit of Fig. 9, $R_1 = R_2 = 1,000$ ohms, $L = 1$ henry, $C = 1 \mu\text{f}$, $\bar{E} = 100$ volts. Find: (a) the time constant of each branch; (b) the maximum value of the current in each branch after the switch is closed; (c) the expression for the current through the switch as a function of time after the closing of the switch. *Ans.:* (a) 0.001 sec, each branch; (b) 0.1 ampere, each branch; (c) 0.1 ampere, independent of time after closing the switch.

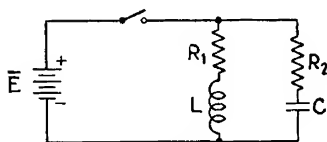


FIG. 9.

10. Figure 10 shows two RC circuits in which $R_1 = R_2 = 10,000$ ohms and $C_1 = C_2 = 10 \mu f$, connected in parallel as shown with the R and C elements in opposite order. The cross branch is a small lamp of very high resistance. Consider that the cross branch has negligible effect on the behavior of the RC circuits, but merely acts as a voltage indicator. Assume that both capacitors are initially uncharged. (a) Describe the behavior of the lamp after the closure of switch S . (b) Sketch approximately to scale the voltage e_{AB} of point A with respect to point B , indicating numerically the initial and final values and also a suitable time scale. When A is at a higher potential than B , e_{AB} is to be considered positive.

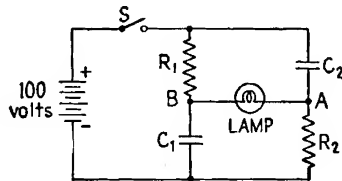


FIG. 10.

Chapter VII

1. The circuit constants for two magnetically coupled circuits are $L_1 = 0.01$ mh, $C_1 = 1,000 \mu\mu\text{f}$, $L_2 = 0.02$ mh, and $C_2 = 1,500 \mu\mu\text{f}$. (a) Find the wavelengths λ_1 and λ_2 of the uncoupled primary and secondary circuits ($M = 0$). (b) The circuits are now coupled, and $M = 0.01$ mh. Find the new resonant wavelengths for either circuit, with the aid of a suitable figure in the text. (c) Check this answer by direct calculation. *Ans.:* (a) 188 m, 326 m; (b) 356 m, 122 m.

2. Find the ratio $\frac{I_2}{I_1}$ for each of the wavelengths of oscillation in Prob. 1. *Ans.:* 2.58, -0.578 .

3. The frequencies of the free oscillation of two identical, magnetically coupled circuits are 490 and 430 keps. Find the coefficient of coupling and the natural frequency of oscillation of each circuit by itself. *Ans.:* 0.1305, 0.650 meps.

4. (a) Prove that the number (m) of oscillations in one beat of the transient oscillation of two freely oscillating, magnetically coupled circuits, tuned individually to the same wavelength, is related to the coefficient of coupling (k) approximately by the equation $k = \frac{4m}{4m^2 + 1}$ (see (3.1), p. 178). (b) Find the coefficient of coupling for the oscillations of Fig. 3.1, p. 178. *Ans.:* (b) 0.0476.

5. In the circuit shown in Fig. 1.1f, p. 176, each $L = 50 \mu\text{h}$, $C_1 = 100 \mu\mu\text{f}$, $C_2 = 200 \mu\mu\text{f}$, and $C_0 = 20 \mu\mu\text{f}$. Find: (a) the coefficient of coupling; (b) the frequency of free oscillation of each circuit alone; and (c) the frequencies of free oscillation of the complete circuit. *Ans.:* (a) 0.123; (b) 2.07 meps, 1.53 meps; (c) 1.507 meps, 2.082 meps.

6. Two identical circuits have the constants $L = 0.01$ mh, $C = 1,000 \mu\mu\text{f}$, $Q = 8$. (a) Find the resistance R of each coil. (b) These circuits are coupled with $k = 0.125$. Using the contour lines of the appropriate figure in the text, draw a curve of the variations of the secondary current as a function of the frequency of the emf applied in the primary circuit. *Ans.:* 100 ohms.

7. In the secondary circuit of the circuit of Prob. 6, C_2 is decreased so as to reduce the maximum secondary current to 0.6 of its former value. Find the new value of C_2 and draw the new resonance curve, indicating the new frequency at which the maximum of the curve occurs. *Ans.:* $564 \mu\mu\text{f}$; 2.22 meps.

8. The primary and the secondary circuits of two magnetically coupled circuits have the following constants: $L_1 = 100 \mu\text{h}$, $R_1 = 5$ ohms, $L_2 = 200 \mu\text{h}$, $R_2 = 4$ ohms. C_1 and C_2 are such that both circuits are tuned individually to an impressed angular frequency $\omega = 2 \cdot 10^6$ radians/sec. What are the values of k and M for: (a) optimum coupling; (b) critical coupling? (c) If E_1 is 0.1 volt, find the voltage across C_2 for (a) and (b). *Ans.:* (a) 0.019, $2.7 \mu\text{h}$; (b) 0.0158, $2.24 \mu\text{h}$; (c) 3.65, 4.47.

9. In Prob. 8 determine the fractional band width for (a) and (b). *Ans.:* 0.020, 0.0247.

10. The coupling network between stages in an i-f amplifier consists of two magnetically coupled circuits as indicated in Fig. 20.1, p. 205, and in Fig. 28.1, p. 406. Four curves of voltage gain are plotted against frequency in Fig. 10. The inductances of the magnetically coupled circuits are identical, and f_0 in Fig. 10 is the proper value of the intermediate frequency. (a) Are the values of Q for the two circuits the same or different? (b) State for each of the four curves of Fig. 10 in which direction (increase or decrease) you would adjust C_1 , C_2 , and M to change the amplification curve to the proper shape. *Ans.:* (a) Q 's are different, $Q_1 < Q_2$; (b) See the table below where zero means no change, plus means increase, and minus means decrease.

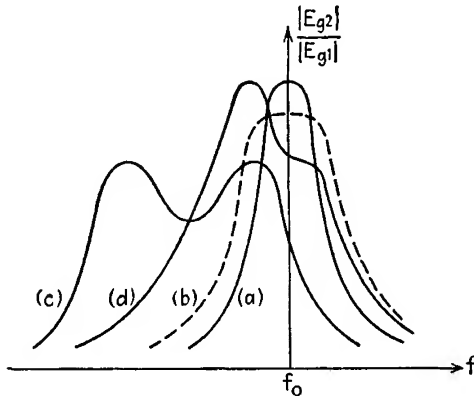


FIG. 10.

TABLE 10

Curve	C_1	C_2	M
a	0	0	+
b	0	0	0
c	—	—	—
d	0	—	0

11. An emf E_1 is impressed in the primary circuit of two magnetically coupled circuits. The primary circuit is tuned to resonance at the frequency of E_1 . (a) How should the coefficient of coupling and the secondary reactance be adjusted to obtain maximum secondary current? (b) How should adjustments be changed to obtain maximum efficiency of power transfer?

12. State in general terms the advantages and disadvantages which result from the use of coupled oscillatory circuits instead of a single oscillatory circuit of like Q .

13. (a) Sketch the curve giving the ratio of the current to the maximum current versus the ratio of frequency to the resonance frequency in a single tuned series circuit, assuming Q to be 100 and independent of frequency and that the magnitude of the applied voltage is constant. (b) On the same dia-

gram sketch the curve of secondary current in two magnetically coupled circuits tuned to the same frequency as the single circuit, assuming that the coupling is critical. All circuits are composed of identical elements and the applied voltage in the two cases is adjusted to give the same maximum values of the two currents. (c) Compare the two values of applied voltage. (d) Show two distinct differences between the curves and explain why and in what application the coupled circuit is particularly useful.

14. (a) Under what conditions of tuning of two magnetically coupled circuits does the system present an impedance of unity power factor to the source of voltage impressed on the primary circuit? (b) Does the system have unity power factor when the adjustments are such as to yield the greatest possible secondary current?

15. In the diagram of Fig. 20.1, p. 205, let the *i-f* transformer be loaded by a resistor R_c' connected across the secondary capacitor C_c . The coupling is adjusted to the optimum value, and C_b and C_c are adjusted for proper alignment. (a) Sketch the response curve as a function of frequency, assuming Q_2 to be one-fourth of Q_1 . (b) The resistor R_c' is now removed, increasing Q_2 by a factor of four, making $Q_1 = Q_2$. Assuming its removal does not detune the circuits, show on the same diagram the response curve that results.

16. The coupled-circuit coupling network shown in Fig. 20.1, p. 205, has the following constants: $R_b = 10$ ohms, $R_c = 40$ ohms, $L_b = L_c = 100$ μ h, $C_c = 100$ μ mf, and $M = 9$ μ h. The system is adjusted to give maximum efficiency of power transfer and to offer a pure resistance in the plate circuit of the first tube. Find (a) the angular frequency of operation; (b) the circuit efficiency; (c) the impedance offered to the plate circuit of the first tube. *Ans.:* (a) 10^7 radians/sec; (b) 95.3 per cent; (c) 4,920 ohms.

17. A generator of E_1 volts and internal resistance of R_0 ohms is connected to a load R_L through a network consisting of two coupled circuits as shown in Fig. 17. Let $R_1 = 5$ ohms, $R_2 = 2$ ohms, $E_1 = 10$ volts, $R_0 = 11$ ohms. (a) Find the values X_1 and X_2 and the smallest value of ωM for maximum power in R_L ; find the maximum power delivered to R_L . Can ωM have other values? Explain. (b) If the system is to be matched at terminals a_1b_1 and also at terminals a_2b_2 , find R_L , ωM , X_1 , X_2 , and the power in R_L . *Ans.:* (a) 0, 0, $8\sqrt{2}$ watts; (b) 4.4 ohms, 6.2 ohms, 0, 0, 0.855 watts.

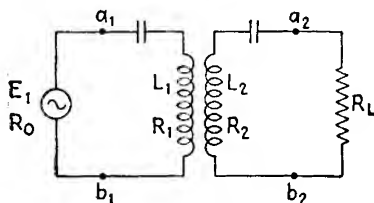


FIG. 17.

18. A generator with an internal resistance of 30 ohms is connected to a load resistance R_L of 50 ohms by a coupling system consisting of two magnetically coupled coils, each in series with a capacitance, the complete network forming two magnetically coupled series circuits. Inductances L_1 and L_2 have resist-

ances such that the Q 's of the coils by themselves are the same and equal to 127.5. The transmission is to have peaks at $f' = 10$ keps and $f'' = 13$ keps. The current in R_L at f' and f'' is the same and equal to twice the current at the minimum transmission between f' and f'' , where f_0 is the frequency at this minimum. The input impedance at f' and at f'' is a pure resistance, the same for both frequencies. The input impedance is also a pure resistance for f_0 . The two circuits are individually resonant at f_0 . The impedance looking backward from R_L is a pure resistance at f' and at f'' . The current in R_L at f' and f'' when $E_1 = 1$ volt is 0.01268 ampere. Find f_0 , coefficient of coupling, L_1 , L_2 , R_1 , R_2 . *Ans.:* 10.75 keps, 0.1065, 1.135 mh, 1.89 mh, 30.6 ohms, 51 ohms.

Chapter VIII

1. Design one section of a resistanceless low-pass filter to cut off at 1,000 cps. The image impedance at zero frequency is to be 500 ohms. *Ans.:* $L = 159$ mh, $C = 0.637$ μ f.

2. (a) How many sections of the filter of Prob. 1 should be connected in cascade in order to secure an attenuation greater than 15 db at a frequency 10 per cent above cutoff? (b) What are the output current and voltage in per cent of input current and voltage under the conditions of (a)? Assume that the filter is terminated in its image impedance. *Ans.:* (a) 2; (b) 17 per cent.

3. Calculate the attenuation α , the phase-lag β , and the characteristic impedance Z_c for one section of the low-pass filter of Prob. 1, at frequencies (a) 500 cps; (b) 2,000 cps. *Ans.:* (a) $\alpha = 0$, $\beta = \pi/3$, $Z_c = 433 + j0$ ohms; (b) $\alpha = 2.63$ nepers, $\beta = \pi$, $Z_c = 0 + j866$ ohms.

4. (a) Design a two-section low-pass filter having a cutoff frequency of 636 cps and working into a constant resistance of 1,000 ohms. Draw a schematic wiring diagram and label each element with its numerical value. (b) What is the attenuation of one section of this filter at a frequency of 1,272 cps? (c) What is the attenuation of the two-section filter at 1,272 cps? What is its phase-lag in radians? *Ans.:* (a) $L = 0.5$ mh, $C = 0.5$ μ f; (b) 2.63 nepers; (c) 5.27 nepers, 2π radians.

5. (a) Design a three-section delay line having a time delay of 30 μ sec and working into a resistance of 600 ohms. Draw a diagram, labeling each element with its numerical value. (b) In what frequency band, approximately, would the above network have a constant time delay? *Ans.:* (a) $L = 6$ mh, $C = 0.0167$ μ f; (b) 16,000 cps.

6. Design a delay line composed of n sections, having a time delay of 50 μ sec, in a frequency band of at least 25,000 cps and working into a resistance of 600 ohms. *Ans.:* $n = 8$, $L = 3.75$ mh, $C = 1.04 \cdot 10^{-8}$ farad.

7. A high-pass filter is composed of shunt inductances $L_2 = 39.8$ mh and series capacitances $C_1 = 0.159$ μ f. Calculate: (a) the cutoff frequency; (b) the image impedance at infinite frequency. *Ans.:* (a) 1,000 cps; (b) 500 ohms.

8. (a) Draw a diagram of a three-section high-pass filter with mid-shunt terminations of the constant- k type, using the smallest possible number of elements. (b) This filter is required to attenuate all frequencies under $f_c = 4,000$ cps and to work into a constant resistance $R = 800$ ohms. Find the numerical value for every element indicated in (a). *Ans.:* (b) 3 capacitors, 0.0248 μ f each; 2 inductors, 31.8 mh each; 2 inductors, 15.9 mh each.

9. You have at your disposal one inductor of 1 mh and two capacitors of 0.1 μ f. (a) Draw the diagrams of the two symmetrical one-section filters which can be built using these three elements. (b) Calculate the limits in cps of the transmission band of the first filter, and the value in ohms of the constant resistance into which it should work. (c) Calculate the same quantities for the second filter. *Ans.:* (b) 0 to 22,500 cps, 70.7 ohms; (c) 11,250 cps to ∞ , 141.4 ohms.

10. Design a constant- k band-pass filter to pass all frequencies between $\omega_1 = 10,000$ and $\omega_2 = 12,100$ radians/sec. The filter is to work into a resistance of 500 ohms. *Ans.*: $L_1 = 476$ mh, $C_1 = 0.0174$ μ f, $L_2 = 4.34$ mh, $C_2 = 1.9$ μ f.

11. Taking the filter of Prob. 1 as prototype, design an m -derived T section which will have infinite attenuation at 1,200 cps. *Ans.*: $m = 0.553$, $L_1' = 88.0$ mh, $L_2' = 49.6$ mh, $C_2 = 0.352$ μ f.

12. Taking the low-pass filter of Prob. 1 as prototype, design an m -derived Π section for the value $m = 0.6$. *Ans.*: (a) $L_1'' = 95.4$ mh, $L_2'' = 42.4$ mh, $C'' = 0.382$ μ f.

13. Using the results of Probs. 1, 11, and 12, draw a diagram of the low-pass composite filter composed of the following networks in cascade: left half of m -derived Π section whose $m = 0.6$; prototype constant- k T section; m -derived T section whose $m = 0.553$; right half of Π section whose $m = 0.6$. Rearrange the diagram in order to use the minimum number of elements. *Ans.*: see Fig. 19.6, p. 234.

Chapter IX

1. A periodic function $e(x)$ has the period 2π and has the values shown in Fig. 1. Find graphically the symmetrical component $C(x)$ and the anti-symmetrical component $S(x)$ of $e(x)$. (b) Write down the Fourier development of $e(x)$. *Ans.:* See Figs. 10.1 and 10.2, p. 245.

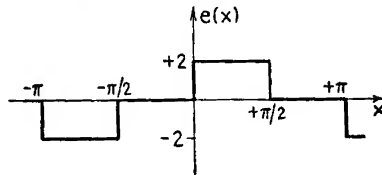


FIG. 1.

2. A periodic voltage of period T sec is zero for three-fourths of the period and in the remaining fourth has the shape of a half-sinusoid of amplitude \hat{E} volts, Fig. 2. Its Fourier development is

$$e(t) = B_0 + B_1 \cos 2\pi \frac{t}{T} + A_1 \sin 2\pi \frac{t}{T} + \dots \quad (2)$$

(a) Obtain the values of the three coefficients B_0 , B_1 , A_1 . (b) Draw a graph of the sum of the three terms of (2). *Ans.:* (a) $B_0 = \hat{E}/(2\pi)$; (b) $-B_1 = A_1 = (2\hat{E})/(3\pi)$.

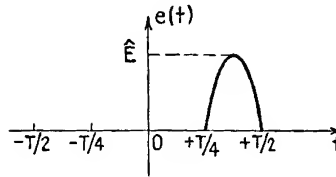
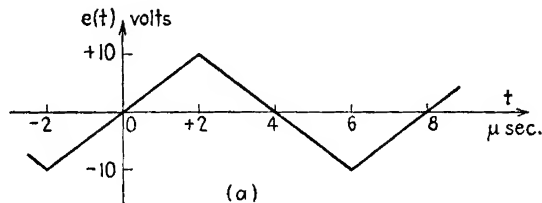


FIG. 2.

3. The voltage e across and the current i through a capacitor are given by Figs. 3a and b. (a) Find the capacitance. (b) Check your result by comparing the fundamental component of each of e and i , their third harmonics, etc. *Ans.:* (a) 1,000 μf .



(a)

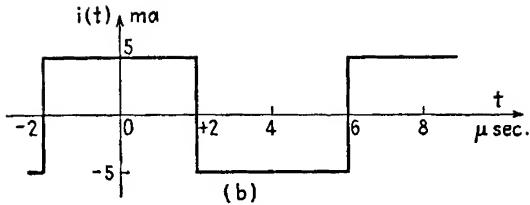


FIG. 3.

4. The input voltage e_1 of a certain network is $e_1(t) = 27 \cos \omega t - 9 \cos 3\omega t$ volts. The network has an attenuation $\alpha = 1$ neper for all frequencies involved and introduces a phase lag of 30° for the fundamental component of e_1 and 90° for the third harmonic. (a) Draw a curve representing e_1 as a function of time, assuming the period of the fundamental component to be 12 milliseconds. (b) Write an expression for the output voltage $e_2(t)$ of the network. (c) Draw a curve representing $e_2(t)$. *Ans.:* (b) $e_2(t) = 8.66 \cos \omega t + 5 \sin \omega t + 3.33 \sin 3\omega t$.

5. A square-wave voltage of frequency f is applied to a circuit having constant L , R , C in series, tuned to the frequency $3f$ and having a Q_r of 100. (a) Find the amplitude of each component of the periodic steady-state current in the circuit. (b) Draw a graph of this current and explain its shape. *Ans.:* (a) Putting $I_3 = (4E)/(3\pi R)$, $I_1 = (9I_3)/800$, $I_5 = (9I_3)/(25 - 9)(100)$, $I_7 = (9I_3)/(49 - 9)(100)$, etc. For $n \gg 1$, $I_n = (9I_3)/(n^2)(100)$.

6. A square-wave voltage has the Fourier development

$$e(t) = \sin \omega t + \frac{1}{3} \sin 3\omega t + \frac{1}{5} \sin 5\omega t + \dots$$

Its period is 1 millisecond. This voltage is applied to a coil of 1 ohm resistance and 2 mh inductance. (a) Write the development of the periodic steady-state current in the coil. (b) Draw a graph of this current and explain its shape.

(c) What are the maximum values of e and i ? *Ans.:* (a) $i(t) \doteq -\frac{1}{4\pi} (\cos \omega t + \frac{1}{3} \cos 3\omega t + \frac{1}{5} \cos 5\omega t + \dots)$; (c) $\pi/4$ volt, $\pi/32$ ampere.

7. (a) What is the frequency of the fundamental component of the voltage shown in Fig. 7? (b) Choose a suitable Y axis and calculate the amplitude of the fundamental component. (c) Make a sketch showing the waveform of Fig. 7 and the fundamental component in the proper phase or time relationship. *Ans.:* (a) 8,333 cps; (b) 63.7 volts.

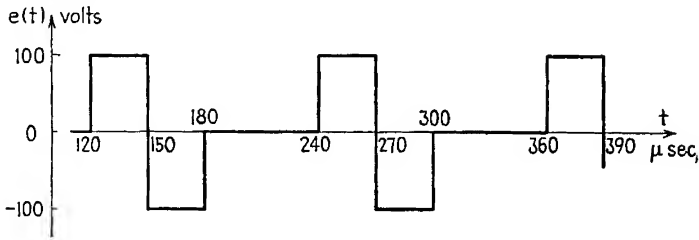


FIG. 7.

8. (a) What are the three lowest frequencies of the harmonic components of the voltage of Fig. 8? (b) The voltage of Fig. 8 can be expressed as the sum

of a symmetrical component $C(t)$ and an antisymmetrical component $S(t)$. Sketch these components to scale. *Ans.:* (a) 125, 375, 625 cps.

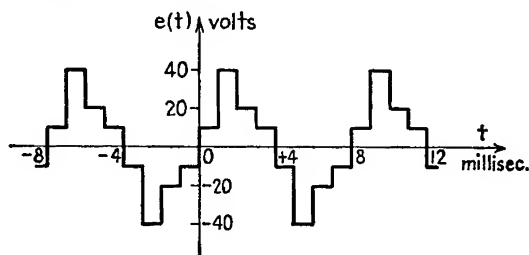


FIG. 8.

9. For the voltage shown in Fig. 9: (a) Determine the d-c component. (b) What are the three lowest frequencies of the alternating components? (c) Make a sketch of this waveform and show thereon an axis locating zero time in such a way that all alternating components can be expressed as cosine terms. (d) Determine the amplitude and phase of the third-harmonic component in accordance with the choice made in (c). *Ans.:* (a) 33.3 volts; (b) 3,333, 6,667, and 13,333 cps; (d) zero.

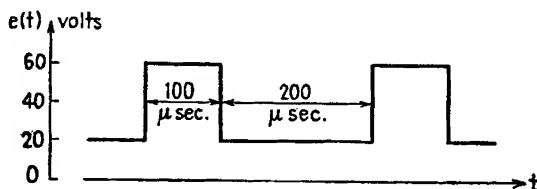


FIG. 9.

10. (a) Determine the d-c component of the voltage of Fig. 10. (b) Make a sketch of the waveform of Fig. 10 and draw thereon the fundamental component, making a guess as to its amplitude, showing it in proper phase or time relationship. (c) Determine the amplitude of the fundamental component [Note: the indefinite integral of $(x \cos x)$ is $(\cos x + x \sin x)$]. *Ans.:* (a) 25 volts; (c) $100(4/\pi^2)$ or 40.5 volts.

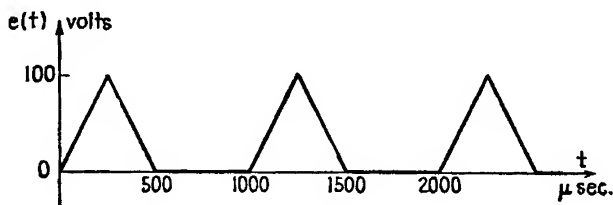


FIG. 10.

Chapter X

1. A diode whose i_b - e_b characteristic is shown in Fig. 1b, is placed in series with a resistance of 25,000 ohms and a battery of unknown but constant potential. A voltmeter V of 50,000 ohms resistance is used to measure the plate voltage of the tube, Fig. 1a. If the voltmeter indicates 75 volts, what is the actual value of the plate voltage when the voltmeter is disconnected? *Ans.*: $e_b \doteq 85$ volts.

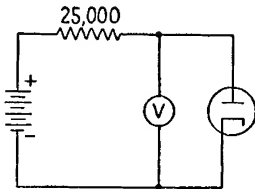


FIG. 1a.

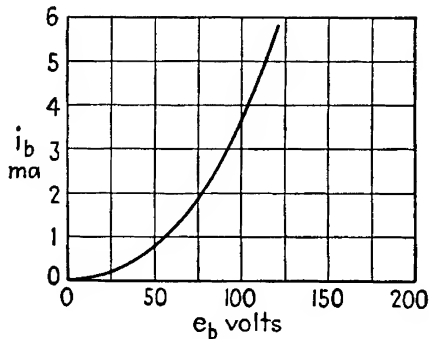


FIG. 1b.

2. Determine the current through the diode in the circuit of Fig. 2a, using the i_b - e_b characteristic of Fig. 2b. The internal resistance of the battery is negligible. *Ans.*: $i_b = 3.0$ ma.

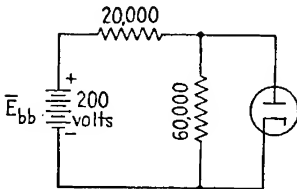


FIG. 2a.

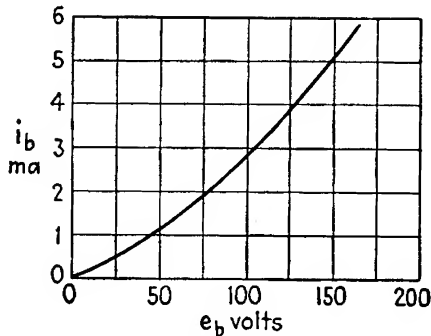


FIG. 2b.

3. The a-c generator in the circuit of Fig. 3a maintains a voltage $e = 70 \sin \omega t$ volts across AB . $R_1 = 10,000$ ohms, $R_2 = 25,000$ ohms. The diode has the characteristic shown in Fig. 3b. (a) At the instant when A is $+70$ volts

with respect to B , what is the potential of C with respect to D ? (b) What is the current through R_2 at this instant? (c) At the instant when A is -70 volts with respect to B , what is the potential of C with respect to D ? (d) Sketch to scale the plate voltage of the diode as a function of the angle ωt . Ans.: (a) 36 volts; (b) 1.44 ma; (c) -50 volts.

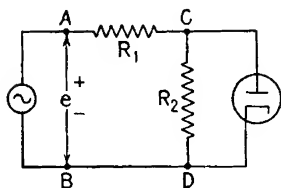


FIG. 3a.

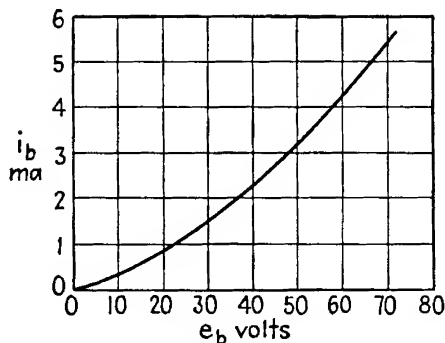


FIG. 3b.

4. Each of the diodes in the circuit of Fig. 4a has the i_b - e_b characteristic shown in Fig. 4b. The sinusoidal voltage applied to the circuit is given by $e = 100 \sin \omega t$. (a) Plot, as a function of the angle ωt , the voltage of point C with respect to point D when the switch S is open. (b) Repeat, with the switch S closed.

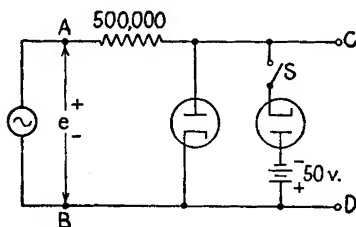


FIG. 4a.

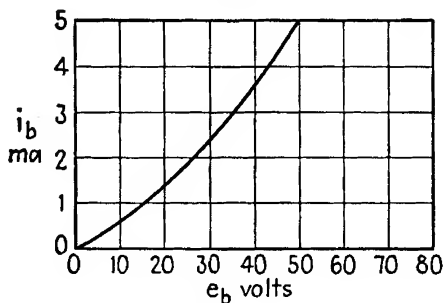


FIG. 4b.

5. A diode having the i_b - e_b characteristic of Fig. 1b, is used in the circuit of Fig. 5. Make a sketch to scale, showing two full cycles of the generator voltage e_{AB} , and two full cycles of the voltage appearing between the terminals C and D . Indicate suitable time and voltage scales. What is the peak current through the diode? *Ans.*: 2.5 ma.

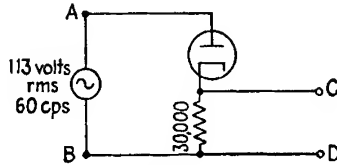


FIG. 5.

6. Each of the diodes indicated in the circuit of Fig. 6 has the characteristic given in Fig. 1b. Make a sketch, approximately to scale, showing the waveform of the current in the resistor R_1 . Indicate numerically the extreme values of the current in R_1 . $R_1 = 12,000$ ohms and $R_2 = 20,000$ ohms.

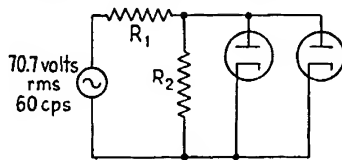


FIG. 6.

Chapter XI

1. A type 2A3 triode is operated as an amplifier with a plate power supply of 350 volts and a plate-load resistor of 2,500 ohms. The grid-bias voltage is -40 volts. Under quiescent conditions, determine: (a) the plate current and the plate voltage; (b) the power output of the plate power supply. A sinusoidal signal of 30 volts amplitude is now applied to the grid. Determine approximately: (c) the maximum and minimum instantaneous values of plate current and voltage; (d) the a-c power delivered to the load; (e) the total power dissipated at the plate of the tube; (f) the plate-circuit efficiency. (g) Is there any evidence of distortion? *Ans.:* (a) $\bar{I}_{b0} = 48$ ma, $\bar{E}_{b0} = 228$ volts; (b) 16.8 watts; (c) 88 ma, 12 ma, 313 volts, 130 volts; (d) 1.7 watts; (e) 9.2 watts; (f) 10 per cent.

2. A triode has the constants $\mu = 18$, $r_p = 12,000$ ohms. A grid bias $\bar{E}_c = -10$ volts and a plate voltage $\bar{E}_b = 250$ volts result in a plate current of 5 ma. (a) Determine the transconductance g_m , assuming μ and r_p to be constant. (b) Draw the lines of constant e_b on the i_b - e_c diagram for $e_b = 214$, 250, 286 volts. (c) Draw the lines of constant e_c in the i_b - e_b diagram for $e_c = -12$, -10 , -8 volts. (d) Draw the lines of constant i_b in the i_b - e_c diagram for $i_b = 2$, 5, 8 ma.

3. One of the triodes of a 6SN7 tube is connected as in Fig. 3. When the switch S is open, the potential of A with respect to ground is $+134$ volts. By graphical means, using the i_b - e_b characteristics of the 6SN7 triode, (a) locate the Q point and determine the amplification factor at the Q point; (b) determine the battery voltage \bar{E}_{bb} . The switch S is now closed. Determine: (c) the change e_c in grid voltage; (d) the potential of A with respect to ground; (e) the change Δe_b in the plate voltage; (f) the voltage amplification. *Ans.:* (f) -13.25 .

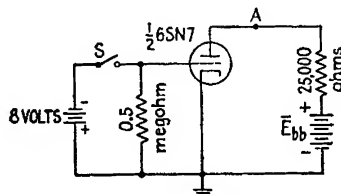


FIG. 3.

4. In the circuit of Fig. 4, $\bar{E}_{b0} = 250$ volts, $\bar{E}_{c0} = -2$ volts, $\bar{I}_{b0} = 0.9$ ma, $r_p = 66,000$ ohms, $\mu = 100$, $C_{gp} = 2.4 \mu\text{f}$, $C_{gk} = 4.0 \mu\text{f}$, $C_{pk} = 3.6 \mu\text{f}$, $E_s = 0.2$ volts rms, $R_g = 20,000$ ohms, $R_L = 150,000$ ohms, and the applied frequency is 0.5 mcps. (a) Compute the magnitude of the voltage amplification, neglecting the interelectrode capacitances. (b) Calculate the grid input capacitance, C_i , assuming that the plate and grid voltages are 180° out of phase. (c) Compute the over-all voltage amplification $|E_p/E_s|$. *Ans.:* (a) $|A| = 69.5$; (b) $C_i = 173 \mu\text{f}$; (c) $|A'| = 6.4$.

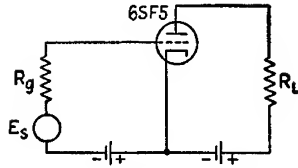


FIG. 4.

5. A 6SJ7 tube is to be used as an amplifier with cathode biasing and a pure resistance load of 50,000 ohms. The quiescent values of grid, plate, and screen voltages are to be -3 , 250, and 100 volts respectively. Under these conditions the plate and screen currents are 3.0 and 0.8 ma respectively; $r_p = 1$ megohm, $g_m = 1,650 \mu\text{mhos}$. (a) Copy the circuit of Fig. 5, completing it by inserting the necessary components for conventional operation. (b) Compute the plate-supply voltage. (c) Calculate the value of each resistor added to the circuit of Fig. 5 and mark the values on the circuit diagram. (d) Calculate the voltage amplification A . (e) Suggest minimum values for screen and cathode by-pass capacitors if the lowest frequency of operation is 60 cps. *Ans.:* (b) 403 volts; (c) 790 ohms, 375,000 ohms; (d) $A = -78.5 + j0$; C_k approx. 50 μf , C_d approx. 0.1 μf .

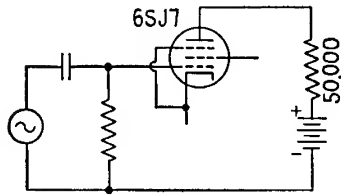


FIG. 5.

6. A single triode of a type 6SN7 tube is used in the circuit of Fig. 6. (a) What is the voltage of the plate with respect to ground when the switch S is open? (b) If S is now closed, will the voltage of the plate increase, decrease, or remain unchanged? If there is any change, what is the new value? *Ans.:* (a) 196 volts; (b) increase, to 249 volts.

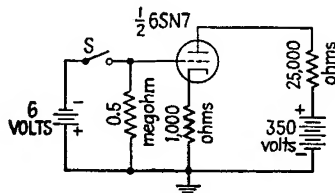


FIG. 6.

7. In Fig. 7, T_1 and T_2 are identical tubes; $e_{s1} = e_{s2} = 2 \sin \omega t$ volts, $R = 5,000$ ohms, $r_p = 10,000$ ohms, and $\mu = 20$ for each tube. Draw the equivalent circuit for each tube and find the numerical expression for i , the instantaneous value of the a-c component of the current through R .

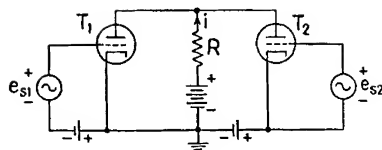


FIG. 7.

8. Using the i_b - e_b characteristic curves, find μ , r_p , and g_m for one triode section of a type 6SN7 tube, at $i_b = 7$ ma and $e_c = -4$ volts.

9. A 6C5 triode is used as an amplifier in the circuit of Fig. 9. The reactance of the plate-load inductor is 10,000 ohms; $r_p = 10,000$ ohms; $\mu = 20$; $e_s = 4 \sin \omega t$; $E_{cc} = 6.0$ volts. (a) Draw the equivalent plate circuit. (b) Find i_p as a function of time. (c) Find e_p as a function of time. Neglect the resistance of the plate-load inductor.

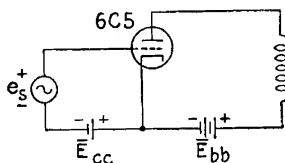


FIG. 9.

10. A 6C5 triode is used in the circuit of Fig. 10. Using $\mu = 20$ and $r_p = 10,000$ ohms, determine the voltage amplification from the input to the output terminals.

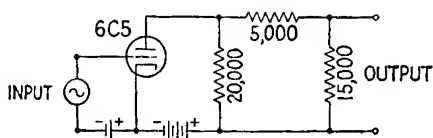


FIG. 10.

Chapter XII

1. The electron beam of a certain cathode-ray oscillograph carries 100 microamperes, and the total accelerating potential is 6,000 volts. The screen material is zinc orthosilicate, having a characteristic phosphor constant A of approximately 2.6 candle power per watt and a "dead voltage" V_0 of approximately 700 volts. Determine the candle power of the fluorescent spot. *Ans.*: Approx. 1.4 candle power.

2. An experimental cathode-ray tube has been constructed, with plane-parallel deflection plates 1.2 cm in length and with 0.3-cm spacing. The screen is 20 cm from the center of the deflection plates. Before entrance into the space between the plates the electron beam has been accelerated by a steady potential difference of 1,200 volts. No additional forward acceleration is imparted to the electron beam after deflection. Neglecting any fringing of the electric field, compute the *electric-deflection factor*, G_E . (Note the comparative insensitivity of this tube as compared with commercially available, triple-anode cathode-ray tubes, which provide for deviation before final acceleration.) *Ans.*: 30 volts/cm.

3. After emerging from the deflection plates of Prob. 2, the deflected electrons appear to have come from a point on the tube axis, in the center of the space between the plates. Can a fluorescent screen of 3-in. diameter be used without interference from a possible electronic shadow, cast on the screen by the trailing edges of the deflection plates? *Ans.*: Yes. Maximum aperture between plate shadows is 10 cm.

4. An electron stream, moving in the x direction with a velocity of $0.2054 \cdot 10^{10}$ cm/sec, is deflected in the y direction by a potential of 60 volts applied across plane-parallel deflection plates, 1.2 cm long and 0.3 cm apart. What is the resultant displacement of the fluorescent spot on a cathode-ray screen, 20 cm from the center of the plates? In the mks system of units the constant A_1 , p. 310, has the value $A_1 = \frac{hL}{d}$. *Ans.*: 2.00 cm.

5. Assuming the electron mass to be constant (since the velocity is relatively low), determine the velocity of an electron stream that has been accelerated by a 6,000-volt difference of potential. *Ans.*: $0.459 \cdot 10^{10}$ cm/sec.

6. An experimental cathode-ray tube has been constructed with internally mounted plane-parallel magnetic poles 1.2 cm in length and with 0.3-cm spacing. The screen is 20 cm from the center of the deflection poles. Before entering the space between the poles, the electron beam has been accelerated by a steady potential difference of 1,200 volts. No additional forward acceleration is imparted to the electron beam after deflection. Neglecting any fringing of the magnetic field, compute the *magnetic-deflection factor*, G_M . *Ans.*: $G_M = 0.049$ weber/square meter/meter.

7. Express the magnetic-deflection sensitivity of this tube in centimeters per gauss. (Note that the magnetic field of the earth, and stray fields of compar-

able intensity, may have an appreciable effect upon such cathode-ray tubes and that magnetic shielding is often required.) *Ans.*: Magnetic-deflection sensitivity is 0.21 cm/gauss for the deflection of electrons.

8. How does the deflection by a magnetic field of a singly ionized nitrogen molecular ion compare with the deflection of an electron by the same magnetic field? *Ans.*: Electron deflection is 225 times greater.

Chapter XIII

1. Three amplifiers are connected in cascade; their complex voltage amplifications are $A_1 = 50/0^\circ$, $A_2 = 40/15^\circ$, and $A_3 = 5/8^\circ$ respectively, the angles being relative phase angles. The input resistance of the first is 80,000 ohms and the load resistance of the third is 8 ohms. (a) Determine the over-all voltage amplification of the cascaded amplifier. (b) Determine the over-all gain (in decibels) of the cascaded amplifier. *Ans.:* (a) $10,000/+23^\circ$; (b) 120 db.

2. In Fig. 6.2, p. 335, the tubes are 6AB7 pentodes operated with the following normal characteristics: $\bar{E}_b = 300$ volts, $\bar{E}_d = 200$ volts, $\bar{E}_c = -3.0$ volts, $\bar{I}_b = 12.5$ ma, $\bar{I}_d = 3.2$ ma, $r_p = 0.7$ megohm, $g_m = 5,000$ μ mhos, $C_i = 8\mu\text{mf}$, and $C_o = 5\mu\text{mf}$. With $R_{b1} = 15,000$ ohms and $R_{c1} = R_{c2} = 0.75$ megohm, the experimentally determined value of the upper half-power frequency of a single stage is 398 kcps ($\omega'' \doteq 2.5 \cdot 10^6$ radians/sec). Determine: (a) the values of R_k , R_d , and \bar{E}_{bb} ; (b) the value of C_1 for a lower half-power frequency of 64 cps ($\omega' = 402$ radians/sec) per stage; (c) the minimum value of C_k for no appreciable cathode degeneration at the lower half-power frequency; (d) the complex and magnitude of voltage amplification of an amplifier consisting of three stages, identical with that of Fig. 6.2, when the frequency of the signal is 86 cps (do not consider the input circuit, C_o , R_{c1}); (e) the lower half-power frequency of the three cascaded stages; (f) the capacitance due to interstage wiring for each stage; (g) the value of R_b for an upper half-power frequency of 2.5 mcps for each stage; (h) the value of L_b (Fig. 18.1, p. 353) for the "compromise" case of shunt-peaking compensation when $f'' = 2.5$ mcps in each stage (it may be assumed that g_m does not change appreciably with the reduced value of R_b); (i) the relative phase angle θ , of the compensated amplifier (three stages) when the applied frequency is 1.5 mcps. *Ans.:* (a) $R_k = 191$ ohms, $R_d = 89,900$ ohms, $\bar{E}_{bb} = 490.5$ volts; (b) $0.0033\mu\text{f}$; (c) $130\mu\text{f}$; (d) $1.93 \cdot 10^5/110.1^\circ$, or $-(0.662 \cdot 10^5 + j1.81 \cdot 10^5)$; (e) 125.5 cps; (f) $14.7\mu\text{mf}$; (g) 2,300 ohms; (h) $64.3\mu\text{h}$; (i) -62.1° .

3. The ideal audio-frequency transformers, T_1 , T_2 , and T_3 , of Fig. 22.1, p. 360, have turns ratios of $\alpha_1 = 40$, $\alpha_2 = 3$, and $\alpha_3 = 0.045$ respectively. The triodes have voltage-amplification factors of 20 and 4 respectively. Determine: (a) the voltage developed across an 8-ohm load in the output circuit when the input voltage is $e_i = 0.1 \sin 2,000\pi t$; (b) the effective plate-load resistance of the output stage. *Ans.:* (a) $e_L = 43.2 \sin 2,000\pi t$; (b) 3,950 ohms.

4. The constants of a typical interstage transformer are given in the example on p. 370. This transformer is used to couple the triode unit of a 6R7 tube to a Class A output stage. The 6R7 is polarized so that $r_p = 8,500$ ohms, $\mu = 16$, and $g_m = 1,900\mu\text{mhos}$. The input capacitance to the output stage is $60\mu\text{mf}$. A 75,000-ohm resistor is connected across the secondary of the transformer. Determine the voltage amplifications of the stage when (a) $\omega = 300$

radians/sec; (b) 6,000 radians/sec; and (c) 34,000 radians/sec. Ans.: (a) $20 + j4$ or $20.4/11.3^\circ$; (b) $20.9/0^\circ$; (c) $18.83/-44.8^\circ$ or $13.39 - j13.30$.

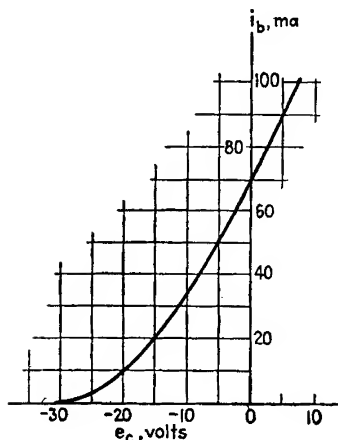


FIG. 5a.

5. (a) The dynamic i_b - e_c characteristic of a triode is given in Fig. 5a. The quiescent grid potential \bar{E}_{c0} is -10 volts. The varying component of the grid potential is $e_g = 10 \sin(2000\pi t + \frac{1}{3}\pi)$. It may be assumed that no harmonics above the second are produced. Determine \bar{I}_{b0} , $(\bar{I}_p)_1$, $(\bar{I}_p)_2$, and the reading of the d-c milliammeter in the plate circuit. (b) A 6F6 power pentode is operated Class A_1 with the following polarizing potentials: $\bar{E}_b = 250$ volts, $\bar{E}_d = 250$ volts, $\bar{E}_c = -16.5$ volts. The plate-load resistance is 7,000 ohms. Assuming the applied signal to be $e_g = 16 \sin 800\pi t$, determine by means of the static characteristics, Fig. 5b: (1) $(\bar{I}_p)_1$; (2) $(\bar{I}_p)_2$; (3) $(\bar{I}_p)_3$; (4) \bar{I}_b ; and (5) \bar{P}_L . Ans.: (a) $\bar{I}_{b0} = 34$ ma, $(\bar{I}_p)_1 = 30.5$ ma, $(\bar{I}_p)_2 = 2.75$ ma, $\bar{I}_b = 36.75$ ma; (b) $(\bar{I}_p)_1 = 31.2$ ma,

$(\bar{I}_p)_2 = 0.25$ ma, $(\bar{I}_p)_3 = -2.67$ ma, $\bar{I}_b = 33.75$ ma, $\bar{P}_L = 3.4$ watts.

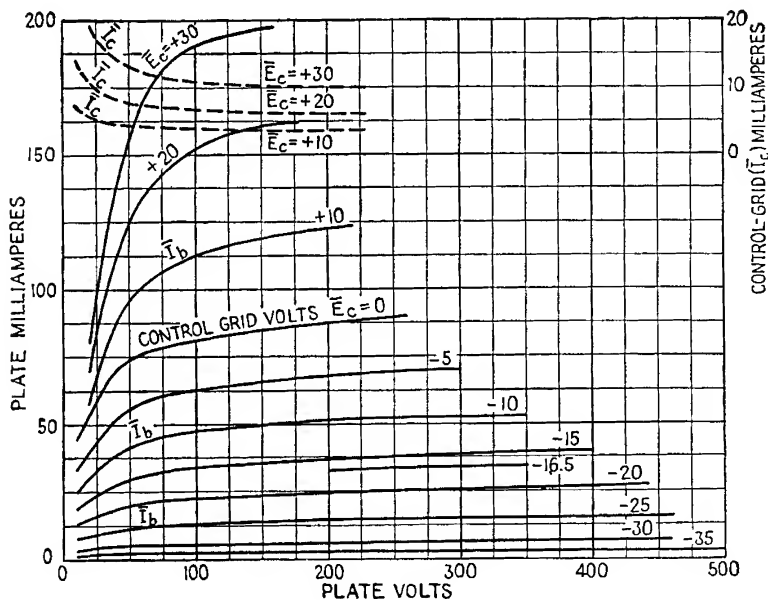


FIG. 5b.

6. Two 2A3 tubes, for which the i_b - e_b chart is given in Fig. 6, are operated in push-pull with $E_{bb} = 300$ volts and $E_c = -60$ volts. The output transformer may be considered ideal. (a) Prepare the composite tube characteristics for the push-pull combination. (b) Locate the Q point of the individual tubes and for the composite tube. (c) Determine graphically the variational plate resistance of the composite tube at the Q point. (d) Assuming no grid current, determine the maximum amplitude of varying component of grid voltage that may be used. (e) Determine the value of R_{pp} (plate-to-plate load resistance) for maximum power output and construct the path of operation. (f) Determine the amplitudes \hat{E}_p and \hat{I}_p for the composite tube and the power delivered to the load when R_{pp} is equal to the value found in (e). (g) Construct the paths of operation for the individual tubes. (h) What is the type of operation of each tube? (i) Replace R_{pp} [as in (f)] by 4,000 ohms and determine \hat{P}_L . *Ans.:* (c) 500 ohms; (d) 60 volts; (e) 2,000 ohms; (f) $\hat{E}_{p \text{ comp}} = 246$ volts, $\hat{I}_{p \text{ comp}} = 126.5$ ma, $\hat{P}_L = 15.55$ watts; (h) Class AB_1 ; (i) 14.3 watts.

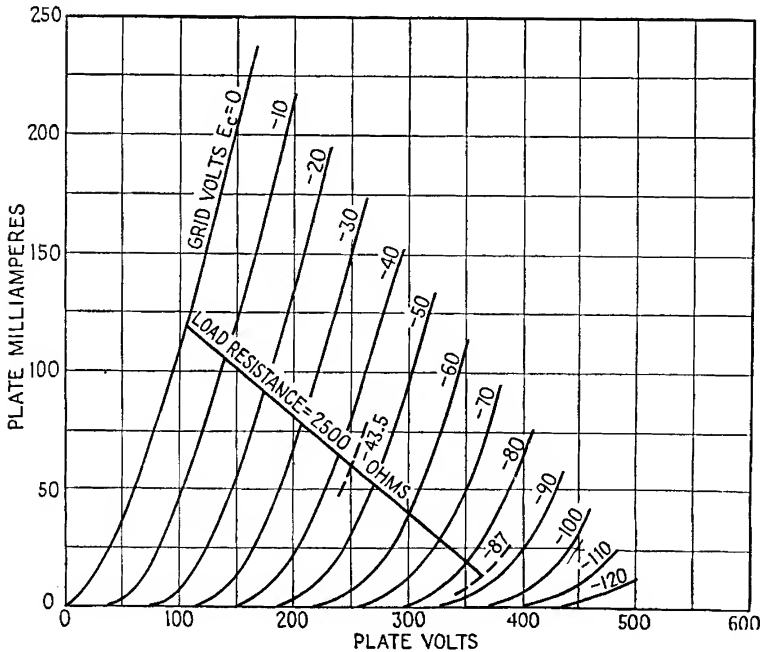


FIG. 6.

7. In the circuit of Fig. 25.1, p. 397, T_1 and T_2 are identical pentodes that are polarized so that $r_p = 10^6$ ohms and $g_m = 5,000 \mu\text{mhos}$. T_3 and T_4 are identical triodes that are polarized so that $r_p = 800$ ohms and $g_m = 1,500 \mu\text{mhos}$. $R_c = R_{c1} = R_{c2} = 0.5$ megohm. $R_{b1} = R_{b2} = 4,000$ ohms. The triodes T_3 and T_4 are connected to a load resistance of 8 ohms by an ideal output transformer, the primary of which is center-tapped. All tubes are operated Class A_1 . Determine: (a) the value of R for the proper operation of the ampli-

fier; (b) the turns ratio $\alpha = N_2/N_1$ of the output transformer for the optimum operation of T_3 and T_4 . *Ans.*: (a) $R = 25,400$ ohms; (b) 0.0707.

8. In Fig. 26.1, p. 399, the pentodes are polarized so that $r_p = 0.75$ megohm and $g_m = 5,000$ μ mhos. The plate circuit is tuned for resonance at 200 mcps at which frequency the Q of the coil L_b is 150 and $C_b = 80$ μ mf. Determine: (a) the value of R_c for a band width of 4 mcps; (b) the voltage amplification of the stage when R_c has the value found in part (a). *Ans.*: (a) 747 ohms; (b) -2.49 .

9. In the TRF stage of Fig. 27.1, p. 402, 6K7 pentodes are used with circuit constants such that $r_p = 1.0$ megohm and $g_m = 1,100$ μ mhos. $L_b = 45$ mh, $R_b = 20$ ohms, $L_c = 300$ μ h, $Q_c = 280$ at 500 keps, 295 at 1,000 keps, and 270 at 1,500 keps. (a) Determine the minimum and maximum values of C_c to cover the frequency range 500–1,500 keps. (b) Assuming the coefficient of coupling k between L_b and L_c to be 0.1, determine the magnitude of the voltage amplification at frequencies of 500, 1,000, and 1,500 keps. (c) Justify any assumptions made in (b). *Ans.*: (a) min 37.5 μ mf, max 337 μ mf; (b) 112.5, 236.9, 325.3.

10. In the amplifier stage of Fig. 28.1, p. 406, the pentodes are polarized so that $r_p = 0.5$ megohm and $g_m = 2,000$ μ mhos. The primary and secondary coils of the transformer are identical, having inductance of 10 mh and $Q = 200$. The primary and secondary circuits are tuned to resonance at 465 keps. Assuming the input resistance to the second tube to be very large, determine: (a) the coefficient of coupling for maximum voltage E_{c2} ; (b) the optimum coupling; (c) the voltage amplification of the stage at resonance for both critical and optimum coupling; (d) the band width of the stage for both critical and optimum coupling. *Ans.*: (a) 0.0178; (b) 0.045; (c) 1,640 for k_c , 1,130 for k_o ; (d) 0.01 for k_c , 0.048 for k_o .

11. The voltage amplification of an amplifier is $10/0^\circ$. If 20 per cent of the output voltage is fed back by means of a resistive negative feedback circuit, what is the actual amplification. *Ans.*: 3.3.

12. An amplifier with equal input and output impedances has a voltage gain of 92.4 db and zero phase shift. If 1 per cent of the output voltage is fed back with a phase shift of 180° , what is the reduced actual amplification obtained, expressed as an absolute ratio? *Ans.*: 100.

13. Negative feedback is applied to an amplifier, which without feedback has $A = 100/0^\circ$. (a) If 1 per cent of the output voltage is fed back negatively by means of a resistive network, what actual gain is obtained? (b) If the leads at one end of the feedback network are switched around so that positive feedback is obtained, does (29.3), p. 412, give the amplification? Interpret this result. *Ans.*: (a) 50; (b) the equation does not apply, since the amplifier oscillates.

14. An amplifier with a voltage amplification of $4 \cdot 10^4$ has a negative feedback circuit that returns 1 per cent of the output voltage (at all frequencies) to the input. Calculate the over-all voltage amplification of the amplifier. *Ans.*: 100.

15. (a) For the circuit of Fig. 15, compute the output impedance Z_{AB} when the potentiometer tap is in the top position. $R = 2,000$ ohms, $R_k = 1,000$ ohms, $g_m = 3,000$ μ mhos. (b) State a use for and the advantages of the circuit of Fig. 15. *Ans.*: (a) 222 ohms.

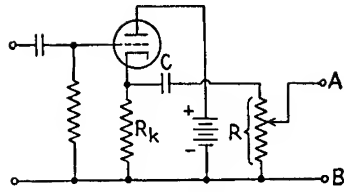


FIG. 15.

16. A loudspeaker with a purely resistive impedance of 10 ohms at 400 cps is to be fed by an output transformer of ratio 15:1. If a pentode must be used as the output tube (a) state briefly two important improvements that can be achieved with negative feedback. (b) Is current-controlled or voltage-controlled feedback to be preferred? Justify your choice. *Ans.:* (a) Less distortion, better matching; (b) voltage control because it reduces the impedance seen by the transformer.

17. Draw a practical circuit for a triode cathode-follower stage in which the direct voltage across the cathode resistor is of the order of 50 volts. (a) State why this circuit is important in video-frequency work. (b) Calculate the voltage amplification and output impedance of the circuit when $g_m = 1,100 \mu\text{mhos}$, $\mu = 100$, and $R_k = 50,000$ ohms. *Ans.:* (b) 884 ohms..

18. An amplifier has a voltage amplification $A = 3,000/0^\circ$. When negative feedback is applied, this voltage amplification is reduced to $A_a = 2,000/0^\circ$. Determine the transmission constant $|\beta|$ of the feedback network. *Ans.:* $|\beta| = 0.000167$.

19. The diagram of Fig. 19 shows a Class A_1 a-f amplifier. The 6J5 triode is polarized so that $\mu = 20$, $g_m = 2,600 \mu\text{mhos}$, $r_p = 7,700$ ohms. The desired plate load impedance for the 6J5 is 20,000 ohms and for the 6V6 is 5,000 ohms. $R_L = 200$ ohms. The transformer T_1 , having a turns ratio $\alpha = 3$, and the transformer T_2 are ideal. The negative-feedback circuit is to be such that a voltage which is equivalent to 10 per cent of E_p for the 6V6 tube is fed back to its grid circuit. (a) Determine the turns ratio of T_2 and indicate the required polarity of its secondary, assuming a positive increment of grid potential to the 6V6 tube. (b) Calculate the value of R_1 and the ratio of R_2 to R_3 . (c) The ratio of E_p to E_g in the 6V6 stage without feedback is $18/180^\circ$. Compute the over-all mid-frequency voltage amplification of the amplifier *with* feedback. *Ans.:* (a) $\alpha = 0.2$; (b) $R_1 = 180,000$ ohms, $R_2 = R_3$; (c) 55.7.

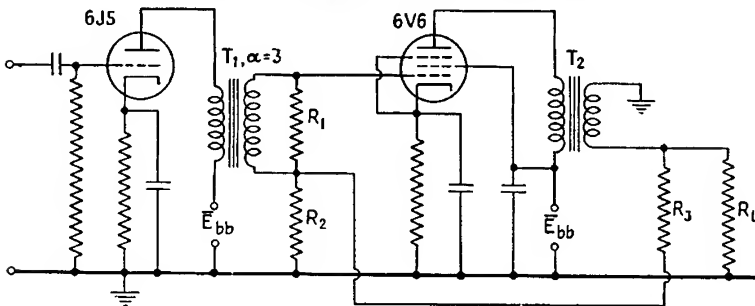


FIG. 19.

Chapter XIV

1. Use the static characteristic curves for the 211 tube given in Fig. 9.1, p. 439, in the solution of this problem. The tube is used as a Class *C* amplifier under the following conditions: $\bar{E}_b = 1,000$ volts, $\bar{E}_c = -150$ volts, $\bar{E}_g = 400$ volts, $\bar{E}_p = 900$ volts. (a) Use the 13-point harmonic analysis to determine \bar{I}_b , $(\bar{I}_p)_1$, \bar{I}_c , and $(\bar{I}_p)_1$. (b) Calculate P_{bb} , P_L , P_p , P_d , P_{ce} , P_g , R_L and η_p . (c) Calculate the constants of the plate tank circuit, assuming $\omega_0 = 5 \cdot 10^6$ radians/sec and $Q_b = 15$. (d) Determine the value of a biasing resistor R_c that would give the assumed bias. *Ans.*: (a) 384 ma, 640, 83.5, 134; (b) 384 watts, 288, 96, 26.8, 12.5, 14.3, 1,410, 75; (c) 0.00213 μ f, $R = 6.23, 18.7 \mu$ h; (d) 1,500.

2. Use the contour curves given in the test for the 211 tube for $\bar{E}_b = 1,000$ volts and $\bar{E}_c = -200$ volts in the solution of this problem. It is desired to operate the tube as a power amplifier to deliver 200 watts to the tank circuit with a plate dissipation of 80 watts. Determine P_d , peak p_d , P_p , \bar{I}_b , \bar{I}_c , $(\bar{I}_p)_1$, $(\bar{I}_p)_1$, η_p , R_c , L , C , R , assuming $Q_b = 10$ and $\omega_0 = 7 \cdot 10^6$ radians/sec. Calculate P_d by Thomas's method and compare with the value obtained from the chart. *Ans.*: 7.7, 66.5, 3.6 watts, 280, 21, 507, 40.5 ma, 71.5 per cent, 22 μ h, 918 μ f, 15.4 ohms.

3. A 211 tube is to be operated at 1,000-volts plate voltage and -200-volts grid voltage. It is desired to obtain the greatest possible power output with the greatest possible efficiency, assuming that the plate dissipation should not exceed 100 watts and that the grid dissipation should not exceed 10 watts. (a) Determine the power output P_L and the plate-circuit efficiency. (b) Determine \bar{I}_b , \bar{I}_c , $(\bar{I}_p)_1$, $(\bar{I}_p)_1$. *Ans.*: (a) 262 watts, 72.4 per cent; (b) 362, 50, 632, 92.5 ma.

4. A 211 tube is used as a Class *C* power amplifier with a plate voltage of 1,000 volts and a grid bias of -200 volts. The characteristics of the tube are given in Figs. 12.3 and 12.4, pp. 447 and 448. The angular frequency of operation is $3 \cdot 10^6$ radians/sec. The tank circuit contains a coil with an inductance of 66.6 μ h and is tuned to give a maximum circuit efficiency of 25/26. The Q of the antenna circuit is 20. The value of \bar{E}_g is 375 volts. Find: C_b , R_L , P_p , P_{bb} , η_p , P_d , peak p_d , and \bar{I}_c . *Ans.*: 0.00255 μ f, 2180 ohms, 50, 230 watts, 78 per cent, 12.5, 102 watts, 37 ma.

5. Given a 211 tube operated at $\bar{E}_b = 1,000$ volts, $\bar{E}_c = -200$ volts. It is assumed that 80 watts is the limit of safe plate dissipation and 10 watts for grid dissipation. Determine the maximum power output, and the corresponding values of P_d , η_p , R_L , and A_p . *Ans.*: 240, 20 watts, 75 per cent, 1500 ohms, 12.

6. Given a 211 tube operated at $\bar{E}_b = 1,000$ volts, $\bar{E}_c = -60$ volts, as a Class *B* amplifier with a plate-load resistance of 3,000 ohms. (a) Determine the maximum excursion of the operating point for a compromise between linear amplification and power output. Determine the carrier point and the values of P_L , P_p , and η_p for no modulation. (b) Using the 9-point analysis, find the average values of P_L , P_p , \bar{I}_b and η_p over the modulation cycle when the

degree of modulation of e_g is 0.7. Determine the percentage modulation of the output voltage for the fundamental modulation frequency. *Ans.:* (a) $\bar{E}_g = 0$ to 200 volts, 49, 115 watts, 29.5 per cent; (b) 62, 104 watts, 166 ma, 37.4 per cent, 68 per cent.

7. (a) Explain fully why the average plate current of a Class C power amplifier is a minimum when the plate circuit is tuned to resonance. (b) Explain under what conditions the average grid current is not a maximum when the plate load is tuned to resonance.

8. A certain Class C power amplifier is operating with an average plate current of 0.40 ampere and a maximum allowable plate dissipation of 1,000 watts. The d-c plate voltage is 10,000 volts and the grid-bias voltage is -300 volts. The plate-load resistance is 10,680 ohms. The average grid current is 0.11 ampere and the amplitude of the grid driving voltage is 1,300 volts. (a) Find: P_d , P_g , η_p , P_L , \bar{E}_p . (b) It is now desired to alter the operation so as to decrease the plate dissipation by approximately 100 watts but to maintain as much power output as possible without increasing the driving power. Show on a contour diagram the paths of operation before and after the change. State whether each of the following quantities is increased or decreased by the change indicated: P_L , \bar{I}_b , η_p , \bar{I}_c , R_L , \bar{E}_v , \bar{E}_p . (c) Calculate the constants of the tank circuit if the loaded Q is to be 12 and $\omega = 5 \cdot 10^6$ radians/sec when the amplifier is operating as in (a). Assume the Q of the coil alone to be 250. *Ans.:* (a) 143, 110 watts, 75 per cent, 3,000 ohms, 8,000 volts; (c) 73.7 watts, 224.7 μmf , 177 μh .

9. The maximum ratings for a certain power tube are: maximum d-c plate voltage = 4,000 volts, maximum d-c plate current = 250 ma, maximum plate dissipation = 300 watts. (a) If the tube is operated as a power amplifier at its maximum ratings, what are the values of maximum power output and plate-circuit efficiency? (b) If the equivalent Q of the loaded tank circuit is to be 12, and the capacitive reactance of the tank capacitor is 200 ohms, what must be the load resistance R_L ? (c) Assuming the Q of the tank coil alone to be 175, what is the circuit efficiency? (d) What effect will increasing the coupling between the tank circuit and the antenna circuit have upon the plate dissipation and plate current? *Ans.:* (a) 700 watts, 70 per cent; (b) 2400 ohms; (c) 93.1 per cent.

10. A Class C power amplifier is connected as in Fig. 14.1, p. 454, except that the tank and antenna circuits are as shown in Fig. 14.2f, p. 455. The grid-bias voltage is supplied from a regulated source. The grid driving voltage is somewhat affected by the amount of driving power required by the amplifier. The system has been properly tuned.

Now assume that because of dry weather the r-f ground resistance increases so that the total antenna resistance is about twice its former value. (a) Indicate on a chart the two positions of the path of operation. (b) Indicate whether each of the following quantities increases or decreases: \bar{I}_b , P_L , P_p , P_{bb} , \bar{I}_c , P_d , η_p .

11. Using the static characteristic curves of the 211 tube, determine the locus of the end point of the path of operation during grid modulation for the following conditions: \bar{E}_g when the modulation is zero, -250 volts; $\bar{E}_b = 1,000$ volts; \bar{E}_g of the a-f modulating frequency, 150 volts; \bar{E}_g of the r-f carrier frequency, 400 volts. The minimum plate voltage at the peak of the modulation

cycle is 100 volts. Determine: the plate-load resistance, the carrier point, the efficiency and P_p at the carrier point.

12. Using the static characteristic curves of the 211 tube, determine the end point of the path of operation for the following plate-polarizing voltages which occur during plate modulation: 100, 300, 500, 750, 1,000, 1,500. The grid-bias potential is -200 volts. When modulation is absent, the steady plate potential is 750 volts and the minimum instantaneous plate potential is 100 volts. Find the plate-load resistance, the peak plate loss, and the peak driving power. Plot the values of \bar{E}_p as ordinates against \bar{E}_b as abscissae, as \bar{E}_b varies during modulation.

13. A Colpitts power oscillator, Fig. 20.2, p. 473, is to operate with the end point of the path of operation at point *a* on the chart for the type 211 tube for $\bar{E}_c = -200$ volts shown in Fig. 12.4, p. 448. The frequency of oscillation is to be 2 mcps. Find the values of the capacitances C_g and C_p , the inductance L_1 of the tank coil, the resistance of the tank coil, and the load resistance which must be in effect coupled in series with the tank coil by the secondary circuit. Assume the Q of the tank coil to be 200 and the effective Q of the loaded tank coil to be 10. Find also the grid-leak or grid-bias resistance R_c . Note: Approximations may be used when the resulting error does not exceed 2 per cent. *Ans.:* $C_g = 575 \mu\text{f}$; $C_p = 256 \mu\text{f}$; $L = 15.8 \mu\text{h}$; $R_{\text{coil}} = 0.992$; $(\omega^2 M^2 / R_2) = 18.8 \text{ ohms}$.

Chapter XV

1. Referring to Fig. 2.1, p. 482, let: $f = 10^6$ cps; $C = 1000 \mu\mu\text{f}$, inclusive of wiring capacitance; $Q = 20$; $E = 1$ volt rms. Assuming the tube to be removed from its socket, what is the resonant value of L and the corresponding current I at resonance? *Ans.:* $L = 25.34 \mu\text{h}$; $I = 125$ ma.

2. Continuing with the same data as in Prob. 1, let: $L_b = 10 \mu\text{h}$; $Q_b = 25$; $\mu = 10$; $r_p = 10,000$ ohms; $|M| = 5 \mu\text{h}$ with regenerative connection. (a) What is the new value of the current I at resonance? (b) How much has the effective resistance been reduced? *Ans.:* (a) $I = 323$ ma; (b) R has decreased from 8.0 ohms to 3.1 ohms.

3. What is the corresponding change in the effective value of L ? *Ans.:* Effective L has increased from $25.34 \mu\text{h}$ to $25.345 \mu\text{h}$.

4. Assuming that no compensating change is made in f or in C , how much has the effective reactance changed? *Ans.:* Effective reactance has increased from 159.22 to 159.25 ohms.

5. Assuming the circuit to be retuned to resonance by a slight change of C , what is the direction and magnitude of this change? *Ans.:* C is reduced from $1,000 \mu\mu\text{f}$ to $999.8 \mu\mu\text{f}$.

6. Assuming the circuit to be readjusted to resonance by changing f , instead of C , what is the direction and magnitude of this change. *Ans.:* f is reduced from 1,000,000 cps to 999,901 cps.

7. At what value of $|M|$ will oscillation commence? *Ans.:* $8.3 \mu\text{h}$.

8. Equation 2.7, p. 483, and Fig. 2.3, p. 485, suggest that a further increase of $|M|$ might cause cessation of oscillation. Is this higher critical value attainable under the particular conditions defined above? *Ans.:* Required value of $|M|$ is $245 \mu\text{h}$, which is *not* physically realizable under the given conditions.

9. Referring to Fig. 8.2b, p. 501, let: $f = 5 \cdot 10^6$ cps; $Q = 250$; $C_1 = 100 \mu\mu\text{f}$; $C_2 = 5 \mu\mu\text{f}$; L be adjusted for resonance. (a) Determine R , assuming $Q = (\omega L)/R$. (b) What voltage across L would be consistent with a 1-volt signal at the terminals $A'B'$? *Ans.:* (a) $R = 26.7$ ohms; (b) $\omega LI = 21$ volts.

10. Referring to p. 513, determine the ranges of oscillation obtainable with variable capacitors, having a maximum capacitance of $1,000 \mu\mu\text{f}$ and an irreducible minimum capacitance of $100 \mu\mu\text{f}$. The following conditions of operation are to be compared: (a) Two such capacitors are to be used in a Wien bridge circuit (or equivalent RC design) in combination with resistors $R_1 = R_2 = 15,915$ ohms. (b) One such capacitor is to be used in a Hartley oscillator (or equivalent LC design) in combination with an inductor $L_1 = 253.3$ mh. *Ans.:* (a) Range of operation without switching, 10,000 cps to 100,000 cps; (b) Range of operation without switching, 10,000 cps to 31,623 cps.

Chapter XVI

1. A voltage-regulator tube is to be used as a pilot rectifier operating with a series load resistor R from a 220-volt rms a-c power source. When the large electrode is the cathode, the glow voltage is constant at 110 volts from 2 ma to 40 ma. When the small electrode is the cathode, the glow voltage increases linearly from 105 volts at 1 ma to 185 volts at 40 ma. (a) What value of R will limit the maximum instantaneous current to 30 ma? (b) With this value of R in series with the tube, what will be the maximum instantaneous current when the small electrode is the cathode? *Ans.:* (a) 6,700 ohms; (b) 23.9 ma.

2. The average grid-control characteristic of a 3C23 thyratron is given by:

Anode voltage	30	35	50	60	100	155	400	800
Grid voltage at start	0	-1	-1.5	-2	-2.5	-3	-4	-5

Plot the grid-control locus for one cycle of a sine wave of voltage on the anode having a peak value of 500 volts. *Ans.:* Graph shows e_c to have the value -4.3 volts at $\theta = \pi/2$.

Chapter XVII

1. A 28-volt battery is to be charged directly from a 110-volt a-c supply by means of a resistor and a single arc-rectifier tube. (a) If the arc drop is 12 volts, what value should R have to limit the peak current to 6 amperes? (b) What is the peak inverse voltage? *Ans.:* (a) $R = 19.25$ ohms; (b) $\hat{E}_{inv} = 183.6$ volts.

2. A half-wave rectifier is to be used to supply direct current to a load consisting of a $4\text{-}\mu\text{f}$ condenser in parallel with a $2,650\text{-ohm}$ resistor. The supply voltage is 550 volts rms at 60 cps. If the tube drop is negligible, (a) what is the average voltage across the load, (b) the average load current, (c) the peak value of the tube current, (d) the peak inverse voltage, and (e) the ripple factor? *Ans.:* (a) $\bar{E}_L = 475$ volts; (b) $\bar{I} = 17.9$ ma; (c) $\hat{i} = 115$ ma; (d) $\hat{E}_{inv} = 1,150$ volts; (e) R.F. = 0.31.

3. A full-wave voltage-doubler circuit is to be used to supply a 280-volt d-c load directly from the 110-volt 60-cps supply. The load resistance is $265,000$ ohms. (a) What value of capacitance should be used to give this average voltage? (b) What will be the ripple factor, (c) the peak tube current, and (d) the peak inverse voltage? *Ans.:* (a) $C = 4\text{ }\mu\text{f}$; (b) R.F. = 0.025; (c) $\hat{E}_{inv} = 295$; (d) $\hat{i} = 87$ ma.

4. Repeat Prob. 3, using a half-wave voltage-doubler circuit. *Ans.:* (a) $\hat{E}_{inv} = 288$ volts; (b) R.F. = 0.01; (c) $\hat{i}_1 = 338$ ma; (d) $\hat{i}_2 = 222$ ma.

5. A full-wave rectifier is to be built to supply a 500-volt 300-ma d-c load with a choke-input filter. The a-c power available is 110 volts rms, 60 cps. What are the voltage and power ratings of the transformer? Neglect tube drop. *Ans.:* $E_S = 555$ volts rms, volt-amperes = 200.

6. If a one-section choke-input filter is to be used between the rectifier and the load of Prob. 5, what size choke should be used? Neglect the resistance of the choke as a first approximation. *Ans.:* $L =$ about 3 henrys (2.95).

7. (a) The full-wave rectifier of Prob. 5 supplies a one-section choke-input filter. (a) What is the rms value of the second harmonic voltage across the capacitor if $L = 4$ henrys and $C = 8\text{ }\mu\text{f}$? (b) What is the second harmonic voltage across the load if a second section is added in which $L = 4$ henrys and $C = 8\text{ }\mu\text{f}$? *Ans.:* (a) 13 volts rms; (b) 0.72 volt rms.

Chapter XIX

1. An r-f current which in the absence of modulation would be given by $i = 10 \sqrt{2} \cos 10^6 \pi t$ amperes is amplitude-modulated 60 per cent by means of a continuous signal having the form $A \sin 6\pi 10^3 t$. (a) Determine f_h , f_l , and m_a . (b) Express the modulated current as numerically as possible in an equation modeled after (2.1), p. 616. (c) Express the modulated current as the sum of several sinusoidal components, in an equation modeled after (3.1), p. 617.

2. Prove by calculation that the modulated current of Prob. 1 would cause an rms ammeter to read 10.87 amperes.

3. Represent the components of the modulated current of Prob. 1 to scale in (a) an amplitude spectrum, (b) an energy spectrum.

4. Draw to scale vectors representing the components of the modulated current of Prob. 1, arranging them like the hands of a clock and in their proper positions at (a) $t = 0$; (b) $t = \frac{1}{12}$ millisecc.

5. An r-f carrier current $i = 10 \sqrt{2} \sin 10^8 \pi t$ amperes is angle-modulated by a continuous signal of the form $A \cos 10^4 \pi t$ to such a degree that the maximum frequency deviation Δf is 12 keps. Determine: (a) the modulation index; (b) the number of appreciable components of the modulated wave and the effective band width of the spectrum. *Ans.*: (a) 2.4; (b) band width 50 keps.

6. What are the answers to the questions of Prob. 5 when the amplitude of the modulating signal is doubled? The modulating equipment remains the same and is assumed to operate ideally. *Ans.*: (a) m_p or $m_f = 4.8$; (b) band width 80 keps.

7. What are the answers to the questions of Prob. 5 when the modulating frequency becomes 3 keps without change in modulating-signal amplitude: (a) if the modulator is an ideally operating frequency modulator; (b) if it is an ideally operating phase modulator? *Ans.*: (a) $m_f = 4$, band width 42 keps; (b) $m_p = 2.4$, band width 30 keps.

8. An amplitude-modulated wave is given by the equation

$$i = \hat{I}(1 + m_1 \sin \omega_1 t + m_2 \sin \omega_2 t) \sin \omega_c t.$$

Determine: (a) the component frequencies of this wave; (b) the amplitude of each component.

9. For the wave of Prob. 8, $m_1 = 0.2$, $m_2 = 0.5$, the carrier-component power is 1 kw, and the load resistance is 100 ohms. Compute: (a) the power due to each component current; (b) the total power of the modulated wave; (c) the maximum possible instantaneous voltage across the load; (d) the rms value of the modulated current. *Ans.*: (a) 1,000, 10, 10, 62.5, 62.5 watts; (b) 1,145 watts; (c) 537 volts; (d) 3.38 amperes.

10. For the current of Prob. 8, with $m_1 = 0.2$, $m_2 = 0.5$, $f_1 = 1,000$ cps, and $f_2 = 3,000$ cps, draw a diagram showing the relative size and position of the vectors representing all of the component currents at the instant when the

vector representing the $\omega_h - \omega_i$ component is 90° ahead of the vector representing the ω_h component. Indicate the values of all angles and the relative direction of rotation of the side-band vectors with respect to the carrier vector.

11. A phase-modulated transmitter is transmitting a signal whose carrier frequency is 25 mcps and whose frequency deviation is $\pm 9,000$ cps when the sinusoidal modulating signal has an amplitude of 1 volt and a frequency of 1,500 cps. (a) Determine the number of side frequencies and the band width of the signal. (b) For the same transmitter, the modulating signal is changed to have an amplitude of 1 volt and a frequency of 3,000 cps. Repeat (a) for these conditions. (c) For the same transmitter, the modulating signal is changed to have an amplitude of 2 volts and a frequency of 1,500 cps. Repeat (a) for these conditions. *Ans.:* (a) 18, band width 27 keps; (b) 18, band width 54 keps; (c) 32, band width 48 keps.

12. A frequency-modulated transmitter is transmitting a signal whose carrier frequency is 25 mcps and whose frequency deviation is $\pm 9,000$ cps when the sinusoidal modulating signal has an amplitude of 1 volt and a frequency of 1,500 cps. (a) Determine the number of side frequencies and the band width of the signal. (b) For the same transmitter, the modulation signal is changed to have an amplitude of 1 volt and a frequency of 3,000 cps. Repeat (a) for these conditions. (c) For the same transmitter, the modulating signal is changed to have an amplitude of 2 volts and a frequency of 1,500 cps. Repeat (a) for these conditions. *Ans.:* (a) 18, band width 27 keps; (b) 12, band width 36 keps; (c) 32, band width 48 keps.

13. An amplitude-modulated current having a carrier frequency of 10^6 cps and 40 per cent sinusoidal modulation at an audio frequency of 10^4 cps is passed through a circuit series-resonant at the carrier frequency. At the carrier frequency the reactance of the coil is 500 ohms and its Q is 100. Assuming R , L , and C to be constant, determine: (a) the percentage modulation of the voltage appearing across the circuit; (b) the phase relationship between the envelope of this voltage and the envelope of the modulated current. *Ans.:* (a) 89.5 per cent; (b) voltage envelope leads current envelope by 63.43° of the audio-frequency cycle.

Chapter XX

1. In a certain plate-modulated Class *C* amplifier represented by Fig. 2.1, p. 649, $\bar{E}_{bb} = 2,000$ volts, $\bar{I}_{bb} = 200$ ma, $e_i = 1,400 \sin 2 \cdot 10^3 \pi t$, $\eta = 0.80$. Assuming idealized operation, determine: (a) m_a ; (b) approximate maximum value of e_b (make any reasonable assumptions); (c) power delivered by \bar{E}_{bb} ; (d) power delivered by T_2 ; (e) the impedance into which T_2 works; (f) r-f output power without modulation; (g) r-f output power with modulation. *Ans.:* (a) 0.7; (b) 6,500 volts; (c) 400 watts; (d) 98 watts; (e) 10,000 ohms; (f) 320 watts; (g) 398.4 watts.

2. Suppose that the oscillator-modulator circuit shown in Fig. 10.1, p. 665, is built using a 6AG7 tube as T_2 and with $L = 0.05$ mh, $C = 13 \mu\text{mf}$, $C_1 = 1 \mu\text{mf}$, $R_1 = 500$ ohms, $\bar{E}_{cc} = -4$ volts, $\bar{E}_{dd} = 150$ volts, and $\bar{E}_b = 200$ volts. (a) Making use of average plate characteristics of the 6AG7 tube shown in a tube handbook, calculate and graph g_m as a function of \bar{E}_{cc} . (b) Calculate an approximate value for the frequency of oscillation without modulation. *Ans.:* (a) $g_m = 8,300 \mu\text{mhos}$ at -4 volts; (b) $f = 5.29$ mcps.

3. Assume that a voltage $e = 0.2 \sin 10^3 \pi t$ is injected in series with \bar{E}_{cc} in the above problem. Calculate approximate values for Δf and m_f . *Ans.:* (a) $\Delta f \doteq 11.5$ keps; $m_f = 23$.

4. Under carrier conditions a plate-modulated Class *C* amplifier is operating as follows: $\bar{E}_{bb} = 2,000$ volts, $E_L = 1,200$ volts rms, plate dissipation 300 watts, plate circuit efficiency 78 per cent. Assuming 50 per cent sinusoidal modulation, determine: (a) d-c plate-circuit power input; (b) a-f plate-circuit power input; (c) plate dissipation; (d) r-f power delivered to the load; (e) maximum instantaneous plate voltage; (f) minimum instantaneous plate voltage; (g) maximum instantaneous current through the d-c plate supply; (h) minimum instantaneous current through the plate supply. Assume ideal operation in the circuit of Fig. 2.1, p. 649. *Ans.:* (a) 1,364 watts; (b) 170.5 watts; (c) 337.5 watts; (d) 1,197 watts; (e) 5,547 volts; (f) 151 volts; (g) 1,023 ma; (h) 341 ma.

5. Derive the results of Row 4 of Table 8.1, p. 661. This principle is used in a laboratory instrument known as a wave analyzer. See *General Radio Company Catalog L*, pp. 130–131, 1948.

6. In the circuit on the extreme left of Fig. 10.3, p. 667, assume that the series impedance of C_2 can be neglected, but that the shunt impedance of C_1 and R_1 cannot. Neglecting interelectrode capacitances, derive an expression for the total input admittance at AB , in terms of C_1 , R_1 , μ , and r_p . Assume that the r-f choke coil has infinite impedance to r-f currents.

7. A certain "frequency-modulation" transmitter employs a phase modulator, a frequency multiplication of 256, and a 75- μsec emphasis. The frequency swing is ± 50 keps at a modulating frequency of 10 keps. (a) What is the phase shift in degrees produced by the modulator stage? (b) The modulating frequency is now reduced to 100 cps and the modulating voltage from the microphone is reduced to one-half the previous value. What is the phase shift

produced by the modulator stage under these conditions? *Ans.:* (a) 1.12° ; (b) 11.6° .

8. In a certain transmitter, phase modulation is produced by shifting the carrier component of an amplitude-modulated wave by an angle of 90° . The percentage of amplitude modulation before the carrier is shifted is 10 per cent and the audio frequency is 800 cps. If the carrier is shifted in phase without change in magnitude and if frequency-multiplier stages with an over-all multiplication of 64 are employed, what is the final frequency deviation? *Ans.:* 5,120 cps.

9. A frequency-modulation transmitter is to be designed after the manner of Fig. 11.1, p. 669. The carrier frequency at the antenna is to be 35 mcps. The frequency deviation for "100 per cent modulation" is to be 25 kcps. The audio-frequency circuits are to be designed to cover only the range from 300 to 3,000 cps, and the transmitter is to be capable of "100 per cent modulation" over this range. No emphasis is used. The phase modulator is capable of producing phase shifts not exceeding 30° without distortion. Determine: (a) the amount of frequency multiplication required; (b) the frequency of the oscillator; (c) the minimum number of frequency-multiplier stages, assuming the highest frequency multiplication per stage to be 4. Note that the over-all frequency multiplication must be the product of integers. *Ans.:* (a) 192; (b) 182.3 kcps; (c) 4.

Chapter XXI

1. A vacuum-tube voltmeter is constructed, following that part of the circuit which is to the right of the input terminals in Fig. 4.3, p. 680. A microammeter is used to measure \bar{I}_R . The scale is calibrated to read rms volts when the voltage impressed has a sinusoidal waveform. (a) A square-wave voltage of 10 volts rms is impressed on the voltmeter terminals. What will be the scale reading? (b) A saw-tooth voltage such as that in Fig. 10.6, p. 246, is impressed. If the scale reading is 10 volts, what is the rms value of the saw-tooth voltage? *Ans.*: (a) 7.07 volts; (b) 8.16 volts.

2. In the circuit of Fig. 6.2, p. 684, assume that the tube is one diode of a type 6H6 tube, $R_1 = 200,000$ ohms, $R_2 = 200,000$ ohms. If the voltage applied to the detector has a carrier value of 10 volts rms and is sinusoidally modulated 50 per cent, what is: (a) the d-c voltage across R_1 ; (b) the rms a-f voltage across R_2 . Neglect the a-f by-passing effect of C_1 and the series a-f impedance of C_2 . *Ans.*: (a) -12.6 volts; (b) 4.0 volts rms.

3. Answer the same questions as in Prob. 2, but with the tube consisting of both of the diodes of the 6H6 tube connected in parallel. *Ans.*: (a) -13.3 volts; (b) 4.3 volts rms.

4. Consider the circuit of Fig. 8.1, p. 691. Assume that the diode is one of the diodes of a type 6H6 tube. $R_8 = 100,000$ ohms, $R_{12} = 1$ megohm, $R_{11} = 50,000$ ohms, $R_{14} = 250,000$ ohms, $R_{16} = 1$ megohm overall. When the r-f voltage applied to the detector circuit has a carrier component of 10 volts rms and 50 per cent sinusoidal modulation, determine: (a) the d-c voltage across C_{15} ; (b) the rms a-f voltage across R_{16} . Assume that the capacitors in the circuit have no adverse effects on the a-f output voltage. *Ans.*: (a) 13 volts; (b) 3.4 volts rms.

5. A diode has the following characteristic: $i_b = 10(e_b + 3)^2 \mu a$. In the circuit of Fig. 12.1, p. 694, $E_{bb} = 6$ volts, and e is a sinusoidal voltage having an amplitude of 2 volts. Determine the following components of the plate current: (a) d-c; (b) rms fundamental a-c; (c) rms second harmonic. *Ans.*: (a) 830; (b) 255; (c) 14.14 μa .

6. A square-law detector has impressed upon it a carrier-frequency voltage having a frequency of 10^6 cps, amplitude-modulated by 2 frequencies, 2,500 and 5,000 cps. What are the component frequencies of: (a) the impressed voltage; (b) the resulting current?

7. A receiver designed for frequency modulation is used to receive a phase-modulated signal. Will the high-frequency tones in the audio-frequency reproducer be too strong or too weak in comparison with the low-frequency tones? Explain. *Ans.*: too strong.

Chapter XXIII

1. A superheterodyne receiver with poor image attenuation is used for the reception of broadcast programs on medium waves, 200 to 590 meters. It has been found that each time the receiver is tuned to a radio station *A* at 1,000 keps, another station *B* at 1,350 keps will also be heard in the loudspeaker, even if the variable band width of the i-f amplifier is given the minimum value, ± 2 keps. Further, it has been observed that if the antenna down-lead is not kept well away from the second detector, a whistle always appears when the receiver is tuned to a third station *C* at 870 keps. (a) Determine the frequency of the whistle. (b) Determine the intermediate frequency of the receiver. (c) Suggest a method to cut down this interference. No extra tube can be added. *Ans.:* (a) 5 keps; (b) 175 keps; (c) Shield the antenna lead, shield and filter the detector output.

2. A superheterodyne for short-wave reception has an intermediate frequency f_i of 1.6 mcps. If the oscillator is tuned above the incoming signal in frequency and the incoming frequency is $f_s = 30$ mcps, (a) determine the oscillator frequency f_o and the image frequency f_{im} . (b) Draw a block diagram of the superheterodyne and indicate the frequencies in each part of the receiver, assuming the station received to be modulated by a 15-keps tone and the band-width control in the i-f amplifier to be set at minimum band width, ± 2 keps. *Ans.:* (a) $f_o = 31.6$ mcps, $f_{im} = 33.2$ mcps; (b) at the r-f input, 30 mcps modulated by 15 keps; at the i-f input, 1.6 mcps modulated by 15 keps; in the remaining part of the receiver, no signal.

3. A superheterodyne receiver is to be designed for use in the 60-mcps band. The intermediate frequency is 3 mcps. Calculate the oscillator frequency and tuning frequency of an image suppression wave trap in the input circuit. *Ans.:* $f_o = 57$ or 63 mcps, $f_{im} = 54$ or 66 mcps.

4. If in a grid-leak detector circuit superregeneration is obtained by means of g_m variation such as $g_m = g_{m0} \sin 2\pi f_q t$: (a) Explain the meaning of f_q . (b) Discuss the limits of the f_q value (hint to (b): $f_{\text{audio}} < f_q < f_{r-f}$).

5. A short-wave superheterodyne employs a triode-hexode as frequency converter. (a) Show by means of a simple diagram the electrode arrangement in the tube for outer-grid injection. (b) Give one reason why the triode-hexode is preferable to a pentagrid converter that employs inner-grid injection. (c) What is the required oscillator frequency if the incoming frequency is 10 mcps and the intermediate frequency 1.5 mcps? *Ans.:* (c) 11.5 or 8.5 mcps.

6. A regenerative grid-leak detector is used for u-h-f reception. (a) Why is it that superregeneration may result from an increase of the grid-leak resistance? (b) How may superregeneration be observed? (Give two indications.) (c) Grade the average superregenerative receiver "good" or "poor" with respect to sensitivity, selectivity, and fidelity.

7. (a) Define what is meant by sensitivity as applied to a radio receiver.

(b) Is the sensitivity a good measure of the distance range that can be covered by a certain receiver or is there any other important factor to be taken into account? (c) Why is the superregenerative receiver more sensitive than the regenerative receiver?

8. The gain in a broadcast-receiver detector of the grid-leak type is increased 1,000 times by regeneration. (a) Why cannot the gain be increased millions of times further by making βA almost equal to 1? (b) In which way does the expression $Q = \omega L / (R - R_r)$ explain side-band cutting? (c) If superregeneration is obtained by varying g_m in the manner given by $g_m = g_{m0} \sin 2\pi f_q t$, explain what f_q stands for and whether there are any limits to the value of f_q . (d) Draw a circuit diagram for a regenerative receiver which without any alteration may also illustrate the case of a superregenerative receiver.

9. Define in a general way what is meant by sensitivity of a radio receiver and suggest a simple setup for the measurement of sensitivity, labeling the various apparatus involved.

10. The heterodyning device shown in Fig. 10 can be operated as (a) a beat-frequency oscillator; (b) a superheterodyne converter; (c) a detector in a superheterodyne code receiver; (d) a frequency multiplier. What types of input voltages are used in each case? Which are the usable output components? How are the undesired frequency components eliminated?

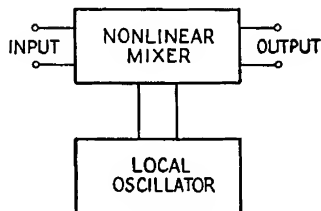


FIG. 10.

INDEX

A

Admittance, 2
 of series or parallel LR or CR circuit, 22
 A-c generator, electrical representation of, 19
 Ammeters, a-c, 721
 d-c, 717
 Amplification, current, 326
 power, 326, 436
 voltage, 325
 at high frequencies, 343
 at low frequencies, 339
 at mid-frequencies, 337
 in multistage amplifier, 326
 Amplification curve, universal, for resistance-coupled stage, 346
 Amplifiers, audio-frequency, 326
 broad-band, 327, 350, 357
 buffer, 494
 cascading, 332
 Class A, 325-422
 Class A_1 , 328
 Class AB , 328
 Class AB_1 , 328
 Class AB_2 , 328
 Class B, 328, 395, 464-466
 Class C, 328, 429-450
 grid-modulated, 466, 655
 plate-modulated, 468, 649
 connections for, 469
 voltages and currents in, 651
 used for modulated signal, 465
 Class C₂, 328
 classification of, 326-328
 band width, 327
 frequency, 326
 single-stage, 328
 current, 326
 current-controlled, 413
 degenerative, 410
 direct-coupled, 332-334

Amplifiers, direct-coupled, Loftin-White, 333
 distortion in, 328-332
 frequency, 329
 nonlinear (harmonic), 328, 371, 375-380, 387, 388
 phase, 331
 feedback in, 410-416
 impedance-coupled, 350
 intermediate-frequency, 327, 405-410, 761
 response curves, measured, 409
 power (*see* Power amplifier)
 push-pull, 382
 equivalent plate circuit for, 393
 radio-frequency, 327
 tuned, 398, 402, 405
 regenerative, 410, 482
 resistance-coupled, 334-360
 transformer-coupled, tuned-primary tuned-secondary, 405
 tuned-secondary, 402
 untuned, 360
 video-frequency, 327, 350, 357
 voltage, 325
 wide-band, 327, 350, 357
 Amplitude, complex, 2, 598
 Amplitude modulation (*see* Modulation, amplitude)
 Amplitude spectra (*see* Spectra, amplitude)
 Analysis, 9-point, 257-259
 13-point, 257-259
 Arc, 522
 negative variational resistance of, 515
 singing, 515
 voltage drop of, 527
 Arc-back, 538
 Arc rectifiers, 526
 Arc tubes, current ratings of, 532
 Arcing, suppression of, 152

Attenuation, 217, 225, 606
 Attenuation band, 209, 218, 226
 Ayrton-Perry winding, 60

B

Balance, to ground, 74
 Band width, 38, 612, 642
 curves of, for k_c and k_o , 203
 fractional, for critical and optimum coupling, 204
 relation of, to frequency deviation, 637
 of tuned circuit, 641, 642, 645, 652
 Bands, side, 617
 Beam-power tube, 301
 Bias, optimum, 373
 Blocking, positive-grid, 477, 492
 Branch currents, 101
 Bridge methods, 84
 Bridged T measurements, 90-92
 Bridges, accuracy of, 84, 85
 a-c, theory of, 79
 balance obtained in, 84
 difference method, 87
 parallel-resonance, 82
 precautions for, at high frequencies, 87
 resonance, 81
 series-resonance, 82
 simple types of, 80

C

Campbell-Foster theorem, 223
 Capacitance, 8, 912
 distributed, 53, 57, 59, 65, 97
 to ground, control of, 89
 input, square-wave measurement of, 175
 interelectrode, 291
 standards of, 85
 wiring, 335, 343, 357-359
 Capacitance load, pure, with rectifiers, 539
 Capacitance-resistance circuit, 138
 applications of, 145
 Capacitor, 7, 912
 air, or gaseous, 67

Capacitor, by-pass, 334
 electrical representation of, 65
 electrolytic, 69
 dry type, 70
 wet type, 70
 liquid-filled, 68
 solid-dielectric, 68
 types of, 67
 Capacitor characteristics, 66
 capacitance, 66
 losses, 66
 maximum voltage rating, 66
 Capacitor-input filters, 570
 Carrier component, disappearance of, 634
 Carrier frequency, stabilization of, 662-663, 668, 673
 Carrier wave, 612-615
 Cascading, 332
 Cascode, 576
 Cathode, 423
 indirectly heated, 264
 virtual, 271
 Cathode bias, 280, 335
 Cathode-bias line, 281
 Cathode degeneration, 336, 414, 416
 Cathode-drop region, 523
 Cathode follower, 417-421
 Cathode-follower circuits, practical, 420
 Cathode-glow (*see* Glow, cathode)
 Cathode-ray beam, 304
 Cathode-ray oscillograph, 320
 Cathode-ray tubes, 304-324
 applications of, 311
 Channel width, 619, 638
 Child's formula, 269, 271
 Choke, swinging, 567
 Choke-input filters, 562
 Circuit constants, equivalent, from measurements, 96
 Circuit efficiency, 206-208
 at maximum power transfer, 208
 Circuit elements, 52-78
 choice of, in amplifiers, 348
 electrical representation of, 76
 measurement of, 79-97
 Circuit response, 21-51

- Circuits, accentuation, 663
 - capacitance-resistance (*see* Capacitance-resistance circuit)
 - counter, 852
 - coupled (*see* Coupled circuits)
 - critical damping of, 156
 - discriminator, 712
 - center-tapped, double-tuned, 713
 - distorter, 663
 - emphasis, 663
 - flip-flop, 843
 - full-wave-voltage-doubler, 550
 - integrator, 663
 - linear sweep, 787
 - multivibrator, 856
 - oscillator (*see* Oscillator circuits)
 - overdamped, 156
 - parallel, 14-19
 - near resonance, 45-48
 - phase-shift, 662, 669
 - phase-splitting, 667, 669
 - reactance-tube, 662, 665, 667
 - representative, 65
 - series, 10-14
 - near resonance, 33-40
 - tank (*see* Tank circuit)
 - thyatron pulse, 799
 - thyatron sweep, 795
 - trigger, 842
 - Eccles-Jordan, 843
 - tuned, as discriminator, 712
 - tuned-grid, 482
 - underdamped, 156
 - vacuum-tube, employing feedback, 726
- Class AB_1 , push-pull operation, 394
 - (*See also* Amplifiers)
- Class B_2 push-pull operation, 396
 - (*See also* Amplifiers)
- Class C , 429-450
 - (*See also* Amplifiers)
- Class C doubler, 480
- Clippers, 813
- Clipping, 687
 - bottom, 689
 - diagonal, 688
- Code reception, 749
- Coils, winding, types of, 57
- Colpitts oscillator, 473
- Compensation, low-frequency, 359
 - shunt-peaking, 352
- Complex quantities, 874
- Composite characteristics, 389-395
- Constant-current generator, 19, 925
- Constant-voltage generator, 19, 925
- Contour curves, for coupled circuits, 190-193
 - for power amplifiers, 447-450
- Control ratio, 529
- Converter, pentagrid (heptode), 758
 - triode, 758
 - triode-heptode, 760
 - triode-hexode, 760
- Coulomb's law, 908
- Coupled circuits, 176-208
 - equivalent impedance of, 183
 - power transfer in, 206
 - response curves of, 198-199, 201
 - two, general representation of, 177
 - types of, 176
 - wavelength of oscillation of, 475
- Coupled systems, parallel-fed, 204
- Coupling, coefficient of, 177
 - critical, 195-196, 408
 - band width for, 203
 - optimum, 200, 409
 - band width for, 202
 - space-charge, 757
 - transitional, 200
- Crystal, grinding of, 502
 - polishing of, 502
 - quartz, 505
- Current, 1
 - circulating, 45
 - peak, 541
 - phase angle of, with respect to voltage, 3
 - resonant rise of, 45
 - secondary, magnitude of, 184
 - maximum, conditions for, 194
 - models for, 186
 - variations in, at optimum coupling, 202
- Current regulator, 578
- Curves, characteristic, 273
 - composite, 389-395
 - (e_b - e_c), 275, 439
 - (i_b - e_b), 274

Curves, (i_b - e_c), 274
 Cutoff, nominal, 586
 Cutoff frequency, 209

D

D, dissipation factor, 17, 65
 Decibels, 325, 870
 Decrement, logarithmic, 158
 Deflecting plates, 308
 Deflection, of beam, by electric field, 308
 by magnetic field, 314
 magnetic, circuit for producing, 811
 voltage source for, 811
 Deflection factor, 310
 Deflection sensitivity, 310, 317
 Degeneration, cathode, 336, 414, 416
 screen-grid, 336
 Deionization time, 530
 Delay, 606
 Delay network, 254
 Demodulation, 674
 Detection, 674-716
 diode, large-signal, 675
 analytical solution, 692
 graphical solution, 682
 of frequency-modulated signals, 712
 grid-circuit, 705
 linear, 675
 in multielement tubes, 705
 plate-circuit, 707
 square-law, 675, 693, 705
 by-passed resistive load, 696
 frequencies involved in, 700
 Detector, 84, 86
 biased, 708
 rectification diagram for, 709
 first, 753
 infinite-impedance, 710
 rectification diagram for, 711
 plate, square-law, 707
 plate-circuit, linear, 708
 power, grid-leak, 706
 regenerative, 741
 second, 753
 superregenerative, 744
 self-quenched, 745
 separately quenched, 745

Detector-circuit impedances, 703
 Deviation ratio, 633
 Dielectric constant, absolute, 913
 relative, 913
 Differentiator circuit, 147, 149, 827
 Diode clipper circuits, 815
 Diodes, 260-271
 current-voltage characteristics of, 265
 gaseous, thermionic, 525
 plane, potential distribution in, 269
 with pure capacitance load, 676
 with *RC* load, 678
 alternative arrangement of, 680
 static characteristics of, 265
 thermionic arc, 526
 thermionic vacuum, 535
 Direct-current source, 925
 constant-current representation, 925
 constant-voltage representation, 925
 Discriminator circuits, 668, 712, 767
 Distortion, 613
 conditions for lack of, 605-606
 delay, 255, 607
 frequency, 255, 329, 607
 harmonic, 329
 high-frequency, 170
 linear, effect of, on symmetrical pulse, 608
 low-frequency, 168
 of modulated wave, 616-617, 619, 640-645
 nonlinear, 328, 376, 607, 613
 phase, 255, 331-332
 types of, 606
 Doubler, frequency, 479
 voltage, 549
 Drag loop, 475, 496
 Drag-loop effect, in crystal vibrator, 504
 Duration, nominal, 586
 Dynatron, 516
 Dynatron oscillator, 517

E

Eccles-Jordan trigger circuit, 843
 Echoes, paired, method of, 609

- Echoes, radio, measurement of, 783
 - Efficiency, plate-circuit, maximum, 433
 - Electric circuit, 915
 - Electric field, 908
 - Electric resistance, 915
 - Electrical discharge, 521
 - Electricity and magnetism, review of, 906-924
 - Electromagnetic induction, 921
 - Electromagnetism, 916
 - Electromotive force, 915-916
 - Electron affinity, 260
 - values of, 261
 - Electron emission, 260-271
 - Electron gun, 306
 - Electronic switches, 835
 - Electrons, electric charge, 907
 - free, in electric field, 911
 - in magnetic field, 920
 - initial velocities of, 269
 - Emission, high-field, 264
 - photoelectric, 264
 - secondary, 264
 - from grid, 477
 - thermionic, 260
 - Emission current, 265
 - Emission efficiency, 263
 - Emitters, oxide-coated, 263
 - pure-metal, 262
 - thoriated-tungsten, 262
 - three types of, comparison of, 263
 - Emphasis curves, 664
 - Energy, from inductor to capacitor, transfer of, 33
 - Energy relations, at series resonance, 32
 - Equivalent networks, 106
- F
- Factor, amplification, 284
 - deflection, 310, 317
 - electric-deflection, 309
 - magnetic-deflection, 316
 - ripple, 543, 558, 564, 571
 - of parallel *RC* load, 544
 - transformer utilization, 559
 - voltage-amplification, 284
 - Fading, selective, 704
 - Feedback, in amplifiers, 410
 - current- and voltage-controlled, 413
 - through impedance of common plate power supply, 421
 - negative, 410
 - output impedance with, 415
 - positive, 410
 - principle of, 411
 - Feedback circuit, 410
 - Feedback factor, 412
 - Figure of merit, 356
 - Filament, heating of, 426
 - Filter circuits, types of, 231
 - H type, 231
 - square type, 231
 - Filtering, r-f, 691
 - Filters, 209-235
 - balanced, 231
 - band-pass, 227
 - design of, 228
 - single, 209
 - band-stop, 229
 - single, 209
 - capacitor-input, 570
 - choke-input, 563
 - composite, 234
 - m*-derived sections of, 232
 - prototype sections of, 232
 - constant-*k*, 224
 - attenuation constant and phase lag of, 225
 - critical inductance, 565
 - decoupling, 422, 749
 - high-pass, 209, 220
 - low-pass, 209
 - attenuation and phase lag of, 216-217
 - characteristic impedance of, 213
 - as delay network, 219
 - input impedance of, 213
 - ladder-type, 212
 - of several identical sections, 219
 - power, 562
 - pulse-shaping, 802
 - resistance-capacitance, 570
 - Π -section, 229
 - Fleming's left-hand rule, 917
 - Fluctuation noise, 774

- Fluorescence, 304
 Focusing, electric-lens method, 305
 magnetic-lens method, 318
 Fourier analysis, 236-259, 556
 harmonic components, amplitude
 and phase of, 241
 C components in, 237-238
 complete, 240
 even components in, 237-238
 numerical method of, 255, 377, 399
 object of, 236
 odd components in, 237-238
 of output signal, 601
 S components in, 237-238
 of symmetrical pulse, 580
 Fourier developments, examples of, 244
 Fourier integrals, 581
 Fourier series, use of, in transmission problems, 249
 Fourier spectrum (see Spectrum, Fourier)
 Fourier transform, 600, 602
 Frequency, division of, by even number, 860
 by odd number, 861
 half-power, 37, 38, 330, 331, 341
 lower, 37
 upper, 37
 instantaneous, 630
 intermediate, 750
 parallel-resonant (see Parallel-resonant frequency)
 side, 617
 Frequency addition and subtraction, 795
 Frequency band, for transmission of signal, 589
 Frequency change, maximum, 632
 Frequency control, automatic (AFC), 767
 Frequency conversion, 753
 Frequency deviation, 632
 Frequency distortion (see Distortion, frequency)
 Frequency division, 792
 Frequency modulation (see Modulation, frequency)
 Frequency multiplication, 479, 794
 Frequency swings, wide, 671
 Frequency-response characteristic, $1/f$, 663
 Frequency-variation method, 95
 Function, antisymmetrical, 238
 circular, 900
 even, 238
 analysis of, 239
 hyperbolic, 900
 odd, 239
 periodic, 236
 symmetrical, 238
 unit, 592
 unit step, 594
 weight, 586
 work, 260

G

 Gain, conversion, 757
 Gain G , 325
 Gas-filled tubes (see Tubes, gas-filled)
 Gases, electrical discharges in, 521
 Generator, a-c, 19
 d-c, 925
 rectangular-wave, 819
 Glow, abnormal, 522
 cathode, 522
 negative, 522
 normal, 522
 Glow discharge, 522
 Glow lamps, 577
 Glow tube, 523
 characteristics of, 524
 rectification by, 525
 symbol for, 523
 g_m , 283
 g_p , 284
 Grid-bias source, power supplied to, 436
 Grid-bias voltage, 333
 Grid-control characteristic, 529
 Grid-control locus, 530
 Grid currents, average component of, 444
 fundamental component of, 444
 maximum, 455
 Grid dissipation, 436

Grid-leak bias, 491
 Grid-voltage supply, 491
 Grids, 425
 decontamination of, 478
 high resistance in series with, 532
 holding, 854
 super-control, 302
 triggering, 854

H

Hairpin winding, 59
 Harmonic, per cent, 377, 379
 Harmonic content of rectified voltages, 556
 Harmonic distortion, cancellation of, 388
 Harmonics, amplitudes of, 377, 379
 even, absence of, 244
 Hay bridge circuits, 82-83
 Henry, 922
 Heterodyning, 753
 High-frequency design procedure, in video amplifiers, 357
 Hum, 778
 Hyperbolic functions, 900

I

$|I_2|$, variation of, with ω , 197
 $|I_2|_{\max}$, conditions for, 194
 value of, 195
 Image interference, 762
 Impedance, 2
 input, of vacuum tube, 291
 general case, 293
 with resistance load, 293
 to transformer, 106
 of two-mesh network, 103
 mesh, 101
 nonlinear, use of, 657
 open- and short-circuit, 107
 output, with feedback, 415
 transfer, 580
 Impedance matching, 98-125
 mechanical analogies in, 124
 Impedance-matching sections, of pure reactance elements, 119-121
 of transmission line, 121

Impedance transformation, 113
 Impulse noise, 775
 Inductance, critical, 565
 mutual, 14, 923
 algebraic sign of, 104
 standards of, 85
 Inductance load, with rectifiers, 543
 Inductive coupling, 104
 Inductor, 9
 electrical representation of, 53
 Inductor characteristics, frequency dependence of, 54
 Insertion loss (*see* Loss, insertion)
 Integration time, 532
 Integrator circuit, 149, 251, 830
 Intensity, range of, 613
 Intensity control, 307
 Interference, 536, 776
 Ionization, 264
 Irradiation, by X rays, 505

K

Keyers, 835
 Kirchhoff equations, 99-100
 Kirchhoff's laws, 916
 in a-c circuit, 12, 14

L

Lattice structures, 230
 LCR circuit, excited by constant-current generator, 40
 excited by constant-voltage generator, 40
 Lens system, 306
 Limiter stage, 772
 Linear elements, 98
 Linearity, improvement of, by use of current limiter, 797
 Lissajous figures, 312
 Litz wire (*see* Litzendraht)
 Litzendraht, 64
 Load, a-c resistance of, 277
 d-c resistance of, 277
 varying, 566
 Load line, audio-frequency, 278, 282, 299, 372, 685
 d-c, 267, 276, 277, 281, 683

Load resistance, optimum, 373, 380
 plate-to-plate, 384
 Load voltage with constant-current
 generator, 25
 Loop resistance (*see* Resistance, loop)
 Loss, insertion, 211
 Losses, effect of, 231
 Low-frequency compensation, 359
 Low-pass filter (*see* Filters, low-pass)

M

Maclaurin's series, 889
 Magnetic core, 58
 Magnetic flux, 917
 Magnetic intensity, 917
 Magnetic particles, powdered, 58
 Magnetomotive force, 384
 Magnetostriction, 505
 Magnetostriction oscillator, 506
 Magnetron, simple, 519
 split-anode, 520
 Mathematics, review of, 865-905
 Mechanical analogues, 136, 137
 of *LRC* circuit, 165
 Mesh currents, 100, 101
 Mesh impedances, 101
 Mhos, 915
 Microphonic noise, 778
 Mixer, crystal, 770
 diode, 770
 vacuum-tube, 760
 (*See also* Tubes, adding)
 Modulating signal (*see* Signal,
 modulating)
 Modulating wave, 612-615, 620-621
 Modulation, absorption processes of,
 648
 amplitude, 615
 methods of, 648
 power during, 619
 angle, 623-640, 661-673
 methods of, 661
 power in, 634
 cathode, 656
 degree of, 616, 652, 659
 distortion, 686
 effect of tuned circuit on, 640

Modulation, frequency, 623, 630
 advantages of, 638
 Armstrong method of, 661
 from phase modulation, 638, 668
 grid, of Class *C* amplifier, 655
 maximum allowable, with diode
 detector, 686
 methods of, 648-673
 of Class *C* amplifier, 649-656
 phase, 623, 625
 with frequency modulation, com-
 parison of, 627
 plate, of Class *C* amplifier, 649-655
 plate-circuit power input, 654
 principles of, 612-647
 reactance-tube, 664
 square-law, 656
 types of, 612
 Modulation index, 633-638, 667, 671-
 673
 Modulator, balanced, 660
 square-law, 657, 659
 Modulator circuit, phase-shift, ex-
 ample of, 669
 Moisture, effects of, 71
 Morse dot, 599
 Motorboating, 421, 749
 μ , 284
 Multielement tubes (*see* Tubes, multi-
 element)
 Multivibrators, 855
 asymmetrical, 858
 basic circuit, 856
 one-cycle, 858

N

Nepers, 217, 872
 Network branch, 99
 Network equations, 99
 Network loop, 99
 Network mesh, 99
 Network notation, 99
 Network theorems, 109
 Networks, 98-125
 compensating, 382, 411
 decoupling, 422, 749
 equivalent, 106
 inverse two-terminal, 223

Networks, linear, 580
 matching, 114
 nondissipative, two-terminal, 222
 phase characteristics and attenuation of, interdependence of, 604
 phase-shifting, 30
 radio-frequency transforming, 114
 two-mesh, input impedance of, 103
 Neutralization, 410, 450-453
 grid, 451
 plate, 451
 Noise, fluctuation, 774, 914
 hum, 778
 impulse, 647, 775
 microphonic, 778
 thermal, 774
 Nondistortion, conditions of, 253
 Nonlinear characteristics, 928
 Nonlinear operation, 281
 Null, in output voltage, conditions
 necessary for, 91
 current, equations at, 93
 relations at, 91-92
 Numbers, complex, 2, 874-878
 conjugate, 877

O

Ohmmeters, 718
 series-type, 718
 shunt-type, 719
 Ohm's law, 9
 Ohms, 915
 Operating point, 275
 Operation, path of, 276, 278, 282, 299,
 372, 391, 440, 442
 Oscillation, forced, 182
 free, frequencies of, 155, 181
 natural frequency of, 484
 frequency of, 155
 intensity of, 485
 intermittent, 492
 transient, 155, 178
 Oscillator circuits, 486
 Colpitts, 473, 488
 Hartley, 487
 Meissner, 488
 tuned-grid, 489
 tuned-grid tuned-plate, 489
 tuned-plate, 489

Oscillators, 471-477, 482-520
 amplitude-modulated, 657
 beat-frequency, 507-508
 blocking, 493, 804-809
 crystal, 499
 dynatron, 517
 electron-coupled, 495
 magnetostriction, 506
 negative-resistance, 514
 phase-shift, 510
 pulsed, 823
 RC, 509
 resistance-stabilized, 513
 relaxation, 149, 413, 784
 transitron, 518
 Wien bridge, 511
 Oscillatory region, 485
 Oscillograph, cathode-ray, 320
 block and circuit diagram of, 321
 wiring diagram of, 322-323
 Output transformers, 382
 Overmodulation, 616, 621, 627, 633
 Overshooting, of capacitor charge, 159

P

Parallel circuits (*see* Circuits, parallel)
 Parallel-fed coupled systems, 204
 Parallel load, of R and C , with rectifiers, 541
 Parallel resonance, 42-51
 with division of resistance between branches, 50
 Parallel-resonant frequency, 43
 ratio of, to series-resonant frequency, 43, 50
 Parallel-resonant impedance, 44
 Parallel-T measurements, 90-93
 Path of operation (*see* Operation, path of)
 Peak current, 541, 543
 Peak voltages, 543
 Peaker circuit, 147, 149, 827
 Pentodes, 295, 299
 Periodogram, 586
 Permeability, 918
 absolute, 918
 relative, 918
 Permittivity, 913

- Phanotron, 527
- Phase angle, relative, 332, 338, 341, 345, 351
- Phase inverters, 397, 798
- Phase lag, 217, 225
- Phase relations, at low frequency, 288
- Phase shift, maximum, in frequency modulation, 633
 - in phase modulation, 626
- Phase shifter, 26
- Phasemeter, 854
- Piezoelectric effects, 499
- Plate circuit, equivalent, 286, 336
 - constant-current form of, 289
 - for push-pull linear operation, 385
 - power relations in, 279, 436
- Plate-circuit efficiency, 436
- Plate conductance, variational, 266, 284
- Plate current, 265
 - average components of, 444
 - fundamental components of, 444
 - minimum, 455
- Plate dissipation, 436
 - P_p , constant value of, contour for, 447
- Plate load, power delivered to, 436
 - sign of R_p for, 295
- Plate resistance, variational, 266, 284
- Plate tank circuit, detuning, effect of, 433
- Plate-voltage supply, 489
- Plates, 424
- Polar form, 2, 875
- Potential, 908
 - ionization, 526
 - polarizing, 335
- Power, average, 4, 6
 - driving, 436
 - instantaneous, 6
 - maximum, conditions for transfer of, 112
- Power amplifier, 326
 - Class A audio-frequency, single-tube, 371-382
 - push-pull, 382-393
 - Class AB push-pull, 393-395
 - Class B, 464
 - push-pull, 395-397
- Power amplifier, Class C, 440-450
 - grid-modulated, 466
 - load circuits for, 455
 - output circuits of, calculation of, 462
 - parallel-feed connections of, 434
 - plate-modulated, 468
 - series-feed connection of, 434
- Power converters, nonsinusoidal, 429
 - sine-wave, 429
 - wide-band, nonsinusoidal operation, 478
- Power-converter tube, current and voltage waveforms for, with resistor as plate load, 431
 - with tuned plate load, 432
- Power filters, 562
- Power output, 375, 493
 - constant value of, contour for, 446
 - maximum, 386
 - maximum undistorted, 373
 - optimum, 373
- Power ratios, 869
- Power supplies, 534-578
 - voltage-regulated, 571
- Power supply, design of, 567
- Power transfer, maximum efficiency of, condition for, 456
- Power tubes, 423-481
 - classification of, 428
 - definition of, 423
 - operation of, 437
 - general principle of, 429
 - narrow-band, 433
 - operation condition, calculation of, 443
 - static characteristic curves of, 438
 - structural features of, 423
 - testing of, 437
- Proximity effect, 64
- Pulse, 580, 647
 - antisymmetrical, analysis of, 594
 - spectra of, 595
 - character of, 590
 - duration of, effect of decreasing, 590
 - instantaneous, 593
 - rectangular, 581
 - periodic, 246

Pulse, symmetrical, Fourier analysis
of, 580
linear distortion on, effect of, 608
spectra of, 587
triangular, periodic, 247
unit, 592
response of network to, 602
voltage, symmetrical, 585
width of spectrum of, 590, 647
Push-pull amplifiers, 382-397

Q

Q, 17, 32, 65, 66
of crystal, 502
Q meter, 93
Q point, 267, 276, 278, 280, 282, 391,
392, 440, 442
 Q_r , 45
determining of, 41
effect of changing, 35
fractional values of, 48
of series branch or loop, 32
 Q_s , 55
Quality factor (*see* Q; Q_r)
Quiescent point (*see* Q point)
Quiescent values, 276
(*See also* Q point)

R

Radio-frequency choke, 57
Radio receivers, 741-782
Radio reception, disturbances to, 773
Range, high-frequency, 337
low-frequency, 337
mid-frequency, 337
Reactance, 11
Reactance-variation method, 95
Reactor, dissipative, equivalent representations of, 16
saturable-core, 831
 R_b and \bar{E}_{bb} , variations in, effects of, 349
Receiver alignment, fundamentals of, 780
Receiver measurements, 778
Receiver servicing, 778
fundamentals of, 780
Receivers, deemphasis in, 664
frequency-modulation, 771
radiation from, 778

Receivers, short-wave, 768
squeals in, 743
tuned-radio-frequency (TRF), 748
Reciprocity theorem, 110
Rectification diagram, 681, 683, 709-711
Rectifier circuits, full-wave, 546
half-wave, 536
polyphase, 553
voltage-doubling, 549
Rectifier constants, 562
Rectifier tubes, typical, characteristics of, 535
Rectifiers, 534-578
arc, waveform of current with, 536
bridge, full-wave, 549
copper-oxide, 721
full-wave, 546
single-phase, output of, 247
waveforms of, 548
glow-tube, 525
half-wave, single-phase, output of, 247
half-wave-voltage-doubling, 551
polyphase, 553
six-phase half-wave, 556
three-phase half-wave, 553
vacuum, 534
Rectigon, 526
Regeneration, 742
Remote-cutoff tube, 302
Residual quantities, 52
Resistance, bleeder, 567
internal, 926
loop, 50
negative, 412
Resistance load, with rectifiers, 536
battery and, 538
Resistance-inductance series circuit, 127
Resistance-variation method, 94
Resistivity, 915
Resistor, 9
composition, 61
electrical representation of, 53
wattage rating of, 61
wire, 59-61
Resistor characteristics, frequency dependence of, 58

- Resonance, parallel, 42-51
 - partial, conditions for, 189, 193
 - series, 31-42
 - Resonance methods, 93
 - Resonant rise, of current, 45
 - of voltage, 45
 - Response, peak, vs. average, 731
 - of simple *LR* and *CR* units, 21
 - Ripple factor (*see* Factor, ripple)
 - RMS value (*see* Value, rms)
- S
- Sawtooth, antisymmetrical, 245
 - symmetrical, 246
 - Schering bridge circuits, 82
 - Screen grid, 295
 - Screen-grid degeneration, 336
 - Screen-grid-tetrode, construction of,
 - employing glass envelope, 296
 - employing metal envelope, 296
 - Self bias, 280, 335
 - Self-inductance, 9, 922
 - Series branch, 32
 - Series circuits (*see* Circuits, series)
 - Series feed, 489
 - Series *LCR* circuit, frequency response
 - of, 33
 - variation of capacitance in, 40
 - Series resonance, 31-42
 - Shielding, 72
 - magnetic, 75
 - Shields, 88
 - Faraday, 76
 - grounded, 75
 - multiple, 74
 - slotted, 76
 - Shot effect, 774
 - Shunt feed, 490
 - Shunt-peaking compensation, 352
 - Signal, amplitude-modulated, 699
 - consisting of finite number of sinusoidal cycles, 596
 - frequency band for transmission of, 589
 - modulating, 612-614
 - for music, 613
 - for spoken words, 613
 - for telegraphy, 613
 - Signal, modulating, for television, 613
 - output, derivation of, from input signal, 602
 - Fourier analysis of, 601
 - smoothing of, 591
 - Signal analysis, 579-611
 - general case, 598
 - Signal duration, nominal, 586
 - Signal generators, 737
 - Signal-to-noise ratio, 779
 - Signals, typical, spectra of, 645
 - Sinusoids, addition of, 3
 - description of, 1
 - Skin depth, 63
 - Skin effect, 56, 62
 - Space charge, 268
 - Spark-suppressor circuit, 151
 - Spectra, amplitude, diagram of, 629
 - of typical signals, 645
 - Spectrum, amplitude, 617-618, 634
 - energy, 589
 - Fourier, 583, 586
 - frequency, continuous, 583
 - Square-cosine, 245
 - Square-sine, 245
 - Square-wave testing, 167
 - Stability, factors affecting, 577
 - Stabilization ratio, 571
 - Stabilizer, amplification-factor bridge
 - type of, 573
 - combination, 575
 - degenerative type of, 574-575
 - internal resistance of, 572
 - transconductance type of, 572
 - Stage, resistance-coupled, polarizing potentials in, 335
 - universal amplification curve for, 346
 - Static characteristics, beam-power tube, 301
 - composite, 391
 - pentode, 299
 - tetrode, 297
 - triode, 274, 439
 - Substitution method, 86
 - Superheterodyne converter circuits, 758
 - Superheterodyne principle, 750

- Superheterodyne receiver, 751
 - duplex-diode-pentode stages in, 765
- Superposition principle, 893
- Superposition theorem, 110
- Superregeneration, 744
- Suppression, carrier and side-band, 704
- Suppressor grid, 299
- Symmetries, quarter-wave, 243
- Synchronization, 788, 791
 - pulse, 791
- T
- T, equivalent, for transformer, 109
- Tank circuit, 114, 434, 454
 - adjustable, 118
 - design of, 115, 454
 - efficiency of, 116
 - filtering action of, 116
 - power in, 436, 653
 - tuning of, 455
- Taylor's series, 888
- Test instruments, 717-740
- Tetrodes, 295-301
 - beam-power, 300
- Theorems, network, 109
- Thermal agitation, 774
- Thermionic emission, 260
- Thermocouple instruments, 723
- Thévenin's theorem, 110
- Thyratron, 529
 - a-c control of, 531
 - screen-grid, 532
- Tickler, 742
- Time constant, 129, 139
- Time delay, 253
- Timing circuits, 783-864
 - applications, simple, 839
- Transconductance, 283
 - conversion, 756
- Transducer, 580
- Transformers, filament, 569
 - ideal, 123-124
 - iron-core, 122
 - modulating, power supplied by, 654
 - output, 382
- Transient oscillation, 155, 178
- Transients, 126-175
 - Transients, occurring with switching, 134
 - oscillatory, 155, 178
 - energy stored during, 163
 - Transition, 518
 - Transmission, side-band, single, 704
 - Transmission band, 209, 218, 225
 - Triggering, 848
 - Triggering controls, 849-850
 - Triggering grids, 854
 - Triode, 272
 - gaseous, thermionic, 529
 - Triode clipper circuit, 816-817
 - Tube coefficients, 283
 - Tube testers, 721
 - Tubes, adding, 837
 - clamping, 823
 - control ratio of, 529
 - converter, 753, 760
 - damping, 821
 - discharge, high-vacuum, 802
 - figures of merit of, 356
 - gas-filled, 521-533
 - operating anodes in parallel, 528
 - glow (*see* Glow tube)
 - hexode-mixer, 302
 - local oscillator, 753
 - mixer, 753, 837
 - multielement, 272-303
 - multipurpose, 301
 - pentagrid-converter, 302, 758
 - pentagrid-mixer, 760
 - power (*see* Power tubes)
 - remote-cutoff, 302
 - sharp-cutoff, 302
 - triode, 273
 - voltage-regulator, 523, 577
 - Tungar, 526
 - Tuning, 454
 - sharpness of, 35
 - Tuning coupled systems, 456-462
 - back-off method, 460
 - convergent, 459
 - correct, 462
 - divergent, 460
 - incremental method, 460
 - Twin T, 93

U

- Units, comparison of, 924
 - Mks system of, 906
- Utilization factor, 559

V

- Vacuum-tube symbols, 282, 931
 - input impedance, 291-295
- Value, average, 3
 - effective, 4
 - peak, 3
 - rms (root-mean-square), 3-4, 6
 - of periodic nonsinusoidal current, 6
- van der Bijl modulator, 656
- Variations, periodic, 1
- Vectors, 873
 - rotating, 2
- Vibration, flectional, 504
 - thickness, 504
 - torsional, 504
- Volt-amperes, reactive, 33
- Voltage, 1
 - control, 788
 - inverse, 538
 - load, average, 542
 - nonperiodic, 579
 - open-circuit, 925
 - peak, 543
 - peak inverse, 538-539
 - phase angle of, with respect to current, 3
 - quench, 744
 - rectified, harmonic content of, 556
 - resonant rise of, 45
 - scanning, 246
 - sinusoidal, 1
- Voltage amplifiers (*see* Amplifiers, voltage)
- Voltage divider, 26
 - compensated, 29, 173
- Voltage ratio, 580
- Voltage system, saw-tooth, 809
- Voltage wave, trapezoidal, 315
- Voltmeters, a-c, 721
 - d-c, 717

Voltmeters, rectifier, 721

- vacuum-tube, 723-737
 - a-c, 727
 - diode, 728-731
 - d-c, 724
 - balanced, 727
 - circuits for balancing out plate current, 725
 - reading, for nonsinusoidal waves, 734
 - typical commercial, characteristics of, 732
- Volume control, automatic (AVC), 690, 764
 - delayed (DAVC), 764

W

- Wagner ground, 88
- Wattless power, 33
- Waveforms, current and voltage, 442
 - determination of, 535
 - Fourier developments of, 244
 - full-wave-rectifier, 547, 548
 - square, 819
- Waves, amplitude-modulated, representation of, by rotating vectors, 621
 - angle-modulated, 629
 - with higher harmonics, decrease in amplitude of, 248
 - modulated, alteration of, by correctly tuned circuit, 640
 - distortion of, 645
 - by off-tune circuit, 645
 - power of, 619-621, 628, 634, 653-654
 - phase-modulated, power of, 628
 - saw-tooth current, 810
 - saw-tooth voltage, 786-797
- Wein bridge circuits, 82, 83
- Willemite, 304
- Winding, Ayrton-Perry, 60
 - bifilar, 59
 - sectionalized, 60
 - fishline, 61
 - resistor, 59
 - reversed in adjacent grooves, 61
- Wire, self-inductance of, 65